NONSTATIONARITIES IN STOCK RETURNS

Cătălin Stărică and Clive Granger*

Abstract—The paper outlines a methodology for analyzing daily stock returns that relinquishes the assumption of global stationarity. Giving up this common working hypothesis reflects our belief that fundamental features of the financial markets are continuously and significantly changing. Our approach approximates the nonstationary data locally by stationary models. The methodology is applied to the S&P 500 series of returns covering a period of over seventy years of market activity. We find most of the dynamics of this time series to be concentrated in shifts of the unconditional variance. The forecasts based on our nonstationary unconditional modeling were found to be superior to those obtained in a stationary long-memory framework and to those based on a stationary Garch(1,1) data-generating process.

I. Introduction

THIS paper reports the results of a nonstationary analysis of the time series properties of daily returns of the S&P 500 index between January 1928 and May 2000: more than seventy years of financial market history. Nonstationary modeling has a long tradition in the econometric literature that focuses on modeling financial returns predating the currently prevalent *stationary, conditional* paradigm [of which processes of the autoregressive conditionally heteroskedastic (ARCH) type and stochastic volatility models are outstanding examples]; see, for example, Officer (1976) or Hsu, Miller, and Wichern (1974). Our endeavor is motivated by growing evidence of instability in the stochastic features of stock returns[1] as well as by an increased awareness of the severe consequences of assuming stationarity when it is not a good modeling approximation (Stărică, 2003; Herzel, Nord, & Stărică, 2004). The project addresses three central questions.

The first question is methodological: How can one analyze index returns in the nonstationary conceptual framework? Our approach is to approximate the nonstationary data locally by stationary models. The changing nature of the probabilistic features of the data—their marginal distribution and dependence structure—imposes a periodic updating of the approximating stationary model. The goal of our methodology is identifying the *intervals of homogeneity,* that is, the intervals where a certain estimated stationary model describes the data well. On an interval of homogeneity the parameters of the data-generating process do not vary much relative to the estimation error of the parameters of the stationary model used as an approximation[2] (see Härdle, Spokoiny, & Teyssière, 2000). The main tool in identifying the homogeneity intervals is a goodness-of-fit test for linear models in the spectral domain, related to the ones proposed in Picard (1985) and Klüppelberg and Mikosch (1996) (see section II of this paper for a detailed description of the methodology, and section III for the relevant statistical results concerning the goodness-of-fit test).

A second, related question is: What type of (major) nonstationarities affect the S&P 500 returns? The in-depth analysis in sections IV to VI as well as the forecasting results in sections VII and VIII indicate the time-varying second unconditional moment as the main source of nonstationarity of returns on the S&P 500 index.[3]

The third question is conceptual: How should we interpret the slow decay of the sample *autocorrelation function* (ACF) of absolute returns? Should we take it at face value, supposing that events that happened a number of years ago have an effect on the present dynamics of returns? Or are the nonstationarities[4] in the returns responsible for its presence? The answer to this question has important implications for estimation and forecasting. In the first case a long history of the time series would carry significant information; in the second case only a short past will be of much use in forecasting. A commonly held belief in the econometric community is that taking the slow decay of the sample ACF at face value (even though it might be caused by nonstationary changes in

---

Received for publication November 19, 2001. Revision accepted for publication September 28, 2004.

* Chalmers University of Technology and Göteborg University; and University of California, San Diego, respectively.

This research has been supported by The Bank of Sweden Tercentenary Foundation and by a grant from Jan Wallanders och Tom Hedelius Stiftelse.

Part of the research presented in this paper was carried out while the first author visited the Department of Economics of the University of California at San Diego. He would like to thank his colleagues there, in particular Rob Engle, for their warm hospitality and generous financial support that made this work possible. The idea of trying the goodness-of-fit test based on the central limit theorem came from Olivier Perrin. The paper gained in clarity due to the careful reading of Holger Drees, of the two referees, and of John Campbell, the associate editor.

[1] See, for example, Stock and Watson (1996). Recently, a growing body of econometric literature (for example, Diebold, 1986; Lamoureux & Lastrapes, 1990; Simonato, 1992; Cai, 1994; Lobato & Savin, 1998; Mikosch & Stărică, 2004) has argued that most of the features of return series that puzzled, through their omnipresence, the so-called stylized facts, including the ARCH effects, the slowly decaying sample ACF for absolute returns, and the IGARCH effect [for definitions and details see Mikosch and Stărică (2003)], are manifestations of nonstationary changes in the dynamic of returns.

[2] For example, if the data-generating process of returns is an independent sequence of random variables with time-changing unconditional variance, then an interval of homogeneity is a period of time when one has reasons to believe that the variance is almost constant (more precisely, that the change in variance cannot be distinguished from estimation error). On the intervals of homogeneity, one approximates the (slowly) changing unconditional variance of returns with a constant. Hence, in the end, the changing pattern of unconditional variance will be approximated by a step function. The resulting model is a process with piecewise constant variance.

[3] Our findings and the modeling methodology that they motivate follow in the steps of Officer (1971) and Hsu et al. (1974). The former, using a nonparametric approach to volatility estimation, reported evidence of a time-varying second moment for the time series of returns on the S&P 500 index and industrial production. The latter modeled the returns as a nonstationary process with discrete shifts in the unconditional variance.

[4] The list of relevant references includes Hidalgo and Robinson (1996), Lobato and Savin (1998), Granger and Hyung (1999), Granger and Teräsvirta (1999), Diebold and Inoue (2001), and Mikosch and Stărică (2004).

The Review of Economics and Statistics, August 2005, 87(3): 503–522

© 2005 by the President and Fellows of Harvard College and the Massachusetts Institute of Technology

the unconditional variance) is a meaningful way of making use of the past in forecasting the future. In other words, estimating long-memory *stationary* models (based on the slow decay of the sample ACF) and using them in forecasting exploits *in a meaningful way* the patterns of change observed in the past. In section VII we investigate the relevance of this assumption by means of a comparison of the forecasting performance of a stationary long-memory process estimated on the series of absolute returns with that of a model based on the paradigm of changing unconditional variance. The results seem to show the superiority of the second method, supporting the hypothesis that the changes of the unconditional variance are the source of long memory in absolute stock returns.

As the paper addresses the issue of volatility and proposes a novel modeling paradigm, a comparison with the Garch framework, the current market leader in volatility modeling, is inevitable. In section VIII we present the results of a forecasting comparison of our methodology with that based on a Garch(1,1) model. We find that the nonstationary unconditional approach produces significantly better volatility forecasts at longer time horizons, between 10 and 250 days.

II. Delimitation of Intervals of Homogeneity

From a traditional time series point of view, the information contained in the time series of daily returns can be split into two components: the sign of the returns and their size. Empirical evidence shows that the sign of daily returns is not predictable. Hence in what follows we will concentrate on studying the time series of absolute returns. More precisely, due to the presence of heavy tails in the absolute returns, we analyze the logarithm of the absolute values of daily returns,[5] $X_t := \log|r_t|$.

In what follows we assume that X_t follow a locally stationary process in the sense of Dahlhaus (1997). In words, we assume that the stochastic features of the data-generating process of the absolute values of daily returns (that is, the marginal distribution and the dependence structure) evolve slowly and smoothly through time as a result of continuous changes in financial markets.

Our methodology consists in *locally* approximating the dynamics of the data X_t with stationary linear models. Assume, for example, that X_t are generated by an ARMA process with parameters that are smooth functions of time:

$$\Phi(t,B)[X_t - \mu(t)] = \Theta(t,B)\,Z_t, \quad Z_t = \sigma(t)\epsilon_t\,, \tag{1}$$

where ϵ_t i.i.d. with $E\epsilon_t = 0$, $E\epsilon_t^2 = 1$, and B denotes the back-shift operator.[6] Our approach yields an approximation of the functions $\Phi(t, B)$, $\Theta(t, B)$, $\mu(t)$, $\sigma^2(t)$ with step functions that are constant on appropriately defined *homogeneity intervals.* This section describes our approach to identifying the homogeneity intervals of the series X_t.

A linear process X_t with unconditional mean μ is defined as

$$X_t - \mu = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j} = \psi(B)\,Z_t, \quad t \in \mathbb{Z}\,, \tag{2}$$

where the innovations (Z_t) are a sequence of i.i.d. random variables with mean 0 and finite variance σ^2, and

$$\psi(z) = \sum_{j=-\infty}^{\infty} \psi_j z^j\,.$$

The assumption

$$\sum_{j=-\infty}^{\infty} |\psi_j| j < \infty$$

ensures that X_t is properly defined as an a.s. absolutely converging series. The linear process (2) has the spectral density function

$$f_\psi(\lambda) = |\psi(e^{-i\lambda})|^2\sigma^2/(2\pi), \quad \lambda \in [0,\pi]\,, \tag{3}$$

and the spectral distribution function

$$F_\psi(\lambda) = \int_0^\lambda f_\psi(x)dx$$

$$= \frac{\sigma^2}{2\pi}\int_0^\lambda |\psi(e^{-ix})|^2 dx, \quad \lambda \in [0,\pi]\,.$$

In the sequel, the linear process (2) with mean μ, noise variance σ^2, and spectral density f_ψ will be compactly denoted by $\mathcal{M}_{\mu,\sigma^2,f\psi}$.

The intervals of homogeneity are constructed by monitoring the changes in the spectral distribution function of X_t as follows.[7] Assume we know that the subsample X_{m_1}, $X_{\mathrm{m}_1+1}, \ldots, X_{m_2}$ is well described by $\mathcal{M}_{\mu,\sigma^2,f_\psi}$, a linear parametric model with mean μ, noise variance σ^2, and spectral density f_ψ. In other words, assume that the interval of homogeneity under construction contains at least the observations m_1 up to m_2. We want to decide whether the following p observations, $X_{m_2+1}, \ldots, X_{m_2+p}$, also belong to the interval. To accomplish this, we test the hypothesis that the linear model $\mathcal{M}_{\mu,\sigma^2,f\psi}$ fits well the subsample $X_{m_2+p-s}, \ldots, X_{m_2+p}$ that contains the p new data points (the size s of the subsample on which the test is conducted is kept constant). A test statistic $T(n, X, \mathcal{M}_{\mu,\sigma^2,f_\psi})$ with a known asymptotic distribution to be specified in the sequel [see equation (6)] is calculated. As the notation emphasizes, the test statistic $T(n, X, \mathcal{M}_{\mu,\sigma^2,f_\psi})$ is a function of the subsample $X_{m_{2+p-s}}, \ldots, X_{m_{2+p}}$ as well as of the model $\mathcal{M}_{\mu,\sigma^2,f_\psi}$. The value obtained is compared with the asymptotic distribution of the test statistic under the null hypothesis that the subsample $X_{m_{2+p-s}}, \ldots, X_{m_{2+p}}$ is a stationary sequence from the model $\mathcal{M}_{\mu,\sigma^2,f_\psi}$. If the value of the statistic falls within the asymptotic confidence interval, the homogeneity interval is extended to include the observations $X_{m_{2+1}}, \ldots, X_{m_{2+p}}$. Otherwise a new homogeneity interval commences with the block $X_{m_{2+p-s}}, \ldots, X_{m_{2+p}}$. The model to describe the data dynamic on the new interval of homogeneity is estimated on this block. The estimated parameters are then assumed to be the true parameters of the new data-generating process, and the procedure of building a homogeneity interval is reiterated.[8]

---

[5] The implications of this analysis for the model choice of the series of returns (r_t) are also discussed in sections V and VI.

[6] This type of stochastic process is a locally stationary process in the sense of Dahlhaus (1997).

[7] The method is related to the one proposed in Picard (1985) for detecting changes in the spectral distribution function of a time series, and further developed for various linear processes, under mild assumptions on the moments of X and the coefficients of the process, by Giraitis and Leipus (1992) and Klüppelberg and Mikosch (1996).

III. A Goodness-of-Fit Test Based on Bartlett's Weighted Integrated Periodogram

We concentrate now on the statistical aspects of the goodness-of-fit test central to the methodology presented above. In this section we define the test statistic $T(n, X, \mathcal{M}_{\mu,\sigma^2,f_\psi})$ and specify its asymptotic distribution. Let

$$\gamma_{n,Y}(h) = \frac{1}{n} \sum_{t=1}^{n-h} (Y_t - \mu)(Y_{t+h} - \mu), \quad h = 0,1,2, \ldots, n-1,$$

denote the *sample autocovariance function* of the stationary sequence Y_t centered at the true mean μ. Recall that the *periodogram* defined as

$$I_{n,Y}(\lambda) = \frac{1}{2\pi} \times \left(\gamma_{n,Y}(0) + 2\sum_{h=1}^{n-1} \gamma_{n,Y}(h) \cos h\lambda\right), \quad \lambda \in [0,\pi], \tag{4}$$

is the natural (method-of-moments) estimator of the spectral density f_Y of the stationary sequence (Y_t); see Brockwell and Davis (1991) or Priestley (1981).

If the linear model $\mathcal{M}_{\mu,\sigma^2,f_\psi}$ defined in equation (2) is the true data-generating process for the subsample $X_{m_{2+p-s}}, \ldots, X_{m_{2+p}}$, the covariances observed in the data ought to match up[9] to the covariances implied by the linear process. The information on the covariance structure of the data is conveniently summarized by the *integrated periodogram* or *empirical spectral distribution function*

$$J_{n,X}(\lambda) := \int_0^\lambda I_{n,X}(y)\, dy = \frac{1}{2\pi} \times \left(\lambda\gamma_{n,X}(0) + 2\sum_{h=1}^{n-1} \gamma_{n,X}(h) \frac{\sin \lambda h}{h}\right) \tag{5}$$

for $\lambda \in [0, \pi]$. Under general conditions, the integrated periodogram is a consistent estimator of the *spectral distribution function* given by

$$F_X(\lambda) = \int_0^\lambda f_X(x)dx, \quad \lambda \in [0,\pi],$$

provided the density f_X is well defined.

The test statistic is defined as

$$T(n,X,\mathcal{M}_{\mu,\sigma^2,f_\psi}) := \sup_{\lambda\in[0,\pi]} \left| \int_{-\pi}^{\lambda} \left(\frac{I_{n,X}(y)}{f_\psi} - \frac{\hat{\sigma}^2}{\sigma^2}\right) dy \right|, \tag{6}$$

where

$$\hat{\sigma}^2 := \int_{-\pi}^{\pi} \frac{I_{n,X}(z)}{|\psi(e^{-iz})|^2}\, dz \tag{7}$$

is an estimate of σ^2. The test statistic is a function of the data, through the periodogram $I_{n,X}\ (y)$, as well as of the hypothesized model $\mathcal{M}_{\mu,\sigma^2,f_\psi}$.

Note that the test statistic (6) does not involve directly the integrated periodogram as defined by equation (5). Instead, Bartlett's weighted form of the integrated periodogram (Bartlett, 1954; cf. Priestley, 1981)

$$J_{n,X,f_\psi}(\lambda) := \int_0^\lambda \frac{\mathrm{I}_{\mathrm{n,x}}(\mathrm{y})}{\mathrm{f}_\psi(\mathrm{y})}\, \mathrm{dy}, \quad \lambda\in[0,\pi], \tag{8}$$

has been used as a building block. Bartlett's weighted form of the integrated periodogram was preferred to the empirical spectral distribution function in constructing the test statistic (6) for the following statistical reason.

Given a finite fourth moment for X and supposing that (X_t) is the linear process (2), the limit of $\sqrt{n}(J_{n,X} - F_\psi)$ in

---

[8] Note that we choose to neglect the estimation error implicit in taking the estimated values of the model's parameters for the real ones. This approximation is acceptable in view of the following two facts. First, a simulation study (see section IV for a brief discussion of the results of the study) indicates that the finite-sample distribution of the test statistic constructed using the true parameters is very close to that of the test statistic obtained using the estimated parameter values. More important, our empirical experience shows that allowing for large deviations of the asymptotic variance of the test statistics from the values prescribed by corollary 3.2 does not change the results of the analysis at all.

[9] If they do not match up, we conclude that a regime shift has occurred and we reestimate the model.

$\mathbb{C}$ [0, π], the space of continuous functions on [0, π] endowed with the uniform topology, is an unfamiliar Gaussian process with a covariance structure that depends on the spectral density f_ψ; see for example Anderson (1993) or Mikosch (1998). Hence, a goodness-of-fit test based on the asymptotic distribution of the integrated periodogram[10] is impractical, as it would require tabulating a distribution for every null hypothesis to be tested.

Dividing the periodogram by the spectral density to produce the test statistic (6) makes the limit process independent of the spectral density. More concretely, the process $J_{n,X,f_\psi}(\lambda)$, properly centered and scaled, converges in distribution in the Skorokhod space $\mathbb{D}$ ([0, π]) to a Brownian bridge; see Shorack and Wellner (1986) as well as theorem 3.1 in the sequel.

Let us recall now the main result, which yields the asymptotic distribution of the test statistic (6).

Theorem 3.1 (Klüppelberg & Mikosch, 1996). Assume that $EZ = 0$, $EZ^4 < \infty$, and denote $\mathrm{var}(Z) = \sigma^2$. Let X_t denote the linear processes (2.2), and $\hat{\sigma}^2$ the estimate of σ^2 defined in equation (7). Then the following holds:

$$\sqrt{n}\int_{-\pi}^{\lambda}\left(\frac{I_{n,X}(y)}{f_\psi(y)} - \frac{\hat{\sigma}^2}{\sigma^2}\right)dy \xrightarrow{d} \pi B(\lambda/\pi)$$

in $\mathbb{D}$ ([0, π]), where the function f_ψ is defined in equation (3), and $B(\cdot)$ is a Brownian bridge.[11]

The following corollary yields the critical values for the hypothesis testing central to the methodology explained above.

Corollary 3.2. Under the hypothesis and with the notation of Theorem 3.1, we have the following:

(a) If X_t is the linear processes (2), then

$$\sqrt{n}\sup_{\lambda\in[0,\pi]}\left|\int_{-\pi}^{\lambda}\left(\frac{I_{n,X}(y)}{f_\psi(y)} - \frac{\hat{\sigma}^2}{\sigma^2}\right)dy\right| \xrightarrow{d} \pi \sup_{\lambda\in[0,\pi]}\left|B(\lambda/\pi)\right|. \quad (9)$$

(b) Denote $\tilde{\sigma}^2 := n^{-1}\sum_{t=1}^{n} Z_t^2$. If $(X_t) = (Z_t)$, then

$$\sqrt{n}\int_{-\pi}^{\lambda}\left(\frac{I_{n,X}(y)}{\sigma^2/2\pi} - \frac{\tilde{\sigma}^2}{\sigma^2}\right)dy \xrightarrow{d} \pi B(\lambda/\pi),$$

$$\sqrt{n}\sup_{\lambda\in[0,\pi]}\left|\int_{-\pi}^{\lambda}\left(\frac{I_{n,X}(y)}{\sigma^2/2\pi} - \frac{\tilde{\sigma}^2}{\sigma^2}\right)dy\right| \xrightarrow{d} \pi \sup_{\lambda\in[0,\pi]}\left|B(\lambda/\pi)\right| \quad (10)$$

in $\mathbb{D}$ ([0, π]).

The distribution of the random variable $\sup_{\lambda\psi[0,\pi]}|B(\lambda)|$ is known.[12]

We would like to emphasize the generality of theorem 3.1, which imposes only low moment restrictions[13] and no distributional restrictions on the innovations Z_t.

We turn now to our detailed data analysis.

IV. Analysis of the Local Dependence Structure

We begin our data analysis with an investigation of the local dependence structure of the logarithm of the absolute returns X_t. The aim of this section is twofold. We want first to measure the strength of the (local) dependence in the log absolute returns, and second to evaluate its time evolution. Toward these goals, we locally estimate AR(1), MA(1), and ARMA(1,1) processes using the methodology outlined in section II. Subsequently, we use the time-varying estimated parameters of these processes to shed light on the two issues of interest.

The data on which our analysis is based are the daily returns of S&P 500 index, $r_t := \log P_t - \log P_{t-1}$, where P_t is the daily closing level of the index between January 3, 1928 and May 25, 2000. There were 390 zeros among 19,261 log returns (2% of the observations), unevenly distributed through the sample. The subsample beginning in 1928 and ending with the introduction of the current definition[14] of the S&R 500 index in March 1957 contains 4% (or 319) zeros, whereas the subsample between 1957 to 2000 contains 0.75% (or 71) zeros. When taking the loga-

---

[10] The null hypothesis of such a goodness-of-fit test is: Data X_t are generated by a linear process (2) with spectral density f_ψ (spectral distribution function F_ψ). The test statistic is (for example) $\sup_{\lambda\in|[0,\pi]}|J_{n,X}(\lambda) - F_\psi(\lambda)|$. Not rejecting the null confirms a good fit of the hypothesized linear model with spectral density f_ψ (spectral distribution function F_ψ) to the sample. Note that the asymptotic distribution of the test statistic changes with the null hypothesis (see also Grenander & Rosenblatt, 1984).

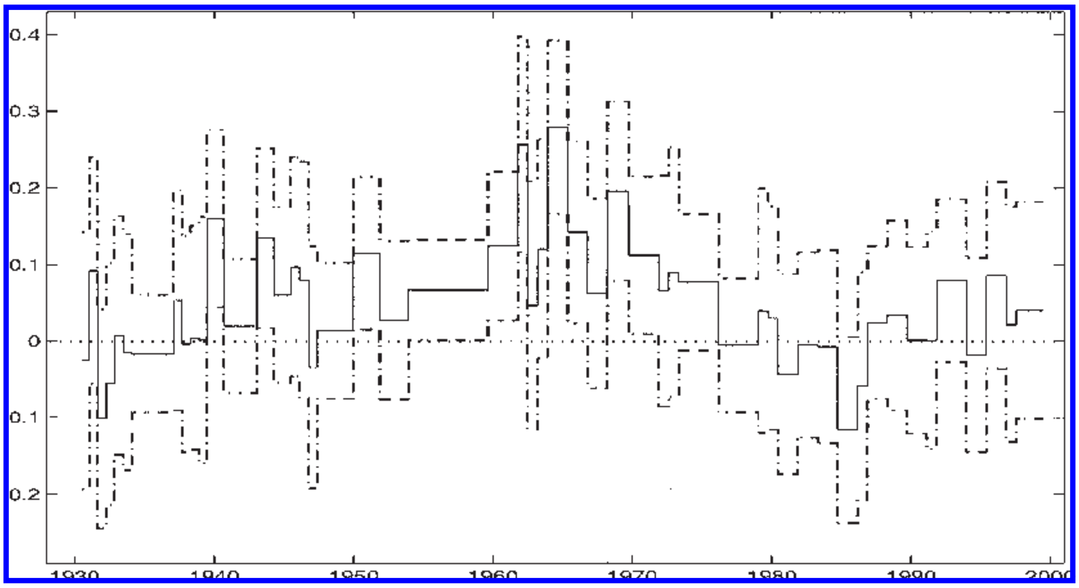
[11] A Brownian bridge on [0,1] is defined as $B(\lambda) := W(\lambda) - \lambda W(1)$ where W is a standard Brownian motion.

[12] Although its distribution function $F(x) = 1 + 2\sum_{k=1}^{\infty}(-1)^k \exp(-2k^2x^2)$ involves an infinite sum, the series is extremely rapidly converging. Usually a few terms suffice for very high accuracy. The limiting distribution was tabulated in Massey (1951, 1952). For example, the 90%, 95%, and 99% quantiles are 1.225, 1.359, and 1.628 respectively.

[13] Although a central limit theorem (CLT) for the integrated periodogram (5) holds true provided $\sigma^2 < \infty$, we are not aware if a similar result holds for Bartlett's weighted periodogram (8). As noted in Klüppelberg and Mikosch (1996), the results in theorem 3.1 are sensitive to large fluctuations in the innovations. This could imply that a goodness-of-fit test based on the statistic (6) has power against an alternative of a linear model with infinite-variance innovations. Because the result in theorem 3.1 fundamentally assumes i.i.d. innovations, we expect the test to have power against alternative hypothesis that violate the i.i.d. assumption of the sequence (Z_t).

[14] In 1957, the S&P 90 was expanded to 500 stocks and became the S&P 500 index. The 500 stocks contained exactly 425 industrials, 25 railroads, and 50 utility firms. This requirement was relaxed in 1988.

FIGURE 1.—THE TIME-VARYING AR COEFFICIENT ϕ_1 (WITH 95% CONFIDENCE INTERVALS) ESTIMATED ON THE INTERVALS OF HOMOGENEITY OF THE LOG ABSOLUTE RETURNS OF THE S&P 500



The intervals of homogeneity correspond to the AR(1) local approximation and are built using the test statistic (6) with $\theta_1 = 0$. Zero is most of the time covered by the interval.

rithm of the absolute returns to produce the sequence $X_t = \log|r_t|$, the zeros have been removed from the sample.[15]

To successfully implement the approach described above we need to specify the statistical elements of the procedure, which we do in the sequel.

A. *The Test Statistic for ARMA(1,1) Processes*

The test statistic we use to build the homogeneity interval is equation (6). For an ARMA(1,1) process defined as

$$\phi(B)(X_t - \mu) = \theta(B)Z_t, \quad \phi(z) = 1 - \phi_1 z, \quad \theta(z) = 1 + \theta_1 z, \tag{11}$$

the filter ψ and the spectral density function are

$$\psi(z) = \frac{\theta(z)}{\phi(z)}, \quad f_\psi(\lambda) = \frac{\sigma^2}{2\pi}\left|\frac{\theta(e^{-i\lambda})}{\psi(e^{-i\lambda})}\right|^2. \tag{12}$$

These quantities, related to the model, together with the data-based periodogram (4), are all we need to construct the test statistic according to equation (6). The method of estimation of the ARMA parameters was that of quasi maximum likelihood.

The size of the subsamples used in the goodness-of-fit test was $s = 250$ days (roughly a business year).[16] The homogeneity interval was extended with 20 observations at a time, that is, $p = 20$ (roughly a business month).

B. *Local Approximation by AR(1) and MA(1) Processes*

We begin by reporting the results of fitting locally stationary AR(1) and MA(1) processes defined by three parameters (μ, σ^2, ϕ_1), (μ, σ^2, θ_1), respectively, to the logarithm of absolute returns. The findings were very similar, and for this reason we concentrate on AR(1) approximation.

Figure 1 displays the time-varying AR(1) coefficient estimated on intervals of homogeneity built using the methodology described in section II. In calculating the test statistic (6), $\theta_1 = 0$ was used in equation (12). One notices that the AR(1) coefficients are, most of the time, not significant. The periods when the coefficient is significant are short, and the AR coefficient is, in absolute value, almost always smaller than 0.18.

The situation is identical when MA(1) processes are used as local approximations, that is, when the test statistic (6) is built setting $\phi_1 = 0$ in equation (12) and MA(1) processes are estimated on the intervals of homogeneity so built: the MA(1) parameter is almost always smaller than 0.15, and most of the time not significantly different from 0. Moreover, the values θ_1 takes are almost identical with those taken by ϕ_1 and displayed in figure 1.[17]

From the graph in figure 1 we conclude that, based on the measures given by the AR(1) and MA(1) local approximations, there is almost no local linear dependence in the log

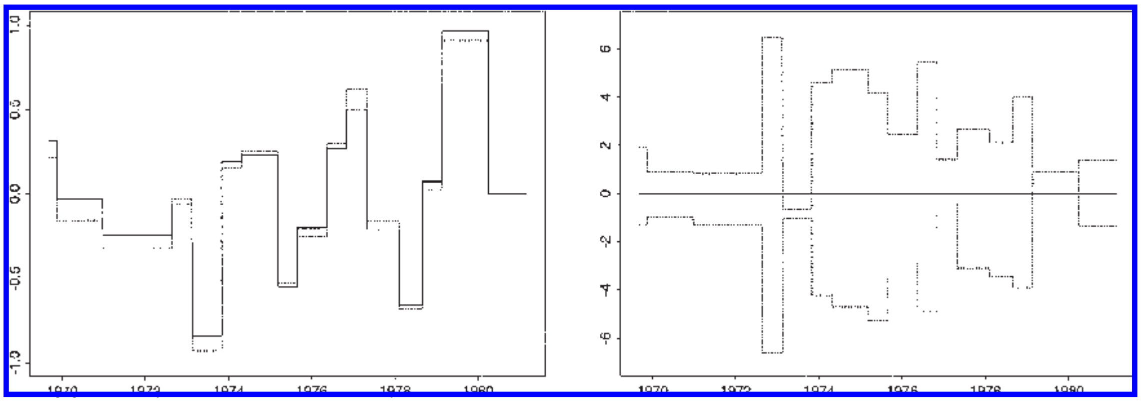
---

[15] The results of the analysis remain unchanged if the zeros are replaced with e^{-6}.

[16] This choice strikes a balance between the need for a sample that is large enough for the asymptotics of theorem 3.1 to work and, at the same time, short enough for the hypothesis of local stationarity to hold. The size s has been chosen empirically. Simulation studies were conducted to evaluate the asymptotic behavior of the test statistic (6).

[17] If the AR(1) [MA(1)] coefficient is not significant, the representation $X_t = \phi_1 X_{t-1} + Z_t$ does not differ from $X_t = \phi_1 Z_{t-1} + Z_t$.

Figure 2.—Left: The Time-Varying AR Coefficient ϕ_1 (*Continuous Line*) and the Minus MA coefficient $-\theta_1$ (*Dotted Line*) Estimated on the Intervals of Homogeneity of the Log Absolute Returns of the S&P500



Right: The 95% Confidence Intervals for the AR Coefficient ϕ_1 Displayed on the Left. Left: The intervals of homogeneity correspond to the ARMA(1,1) local approximation scheme. Right: Zero is covered by the interval most of the time.

absolute returns. Besides short episodes when the linear dependence is low, the data are uncorrelated.

C. *Local Approximation by ARMA(1,1) Processes*

The overall picture of local dependence does not change when one uses ARMA(1,1) processes as local stationary approximations of the dynamics of log absolute returns. Figure 2 displays AR and MA coefficients (together with confidence intervals) estimated on intervals of homogeneity defined using a test statistic (6) corresponding to an ARMA(1,1) process. For the sake of visual clarity, the figure only displays the decade 1970–1980.[18]

Figure 2 summarizes two remarkable findings of the local dependence analysis based on approximating the true data-generating process with ARMA(1,1) stationary processes. *First,* the estimated AR coefficient ϕ_1 and the minus MA coefficient $-\theta_1$, although taking a wide range of values, are always very close. This situation corresponds to the following particular form of equation (11):

$$(1 - \phi_1 B)(X_t - \mu) = (1 - \phi_1 B) Z_t,$$
$$\text{that is,}\quad X_t - \mu = Z_t\,.$$

Second, the 95% confidence intervals contain 0 most of the time. This situation is typical of fitting ARMA(1,1) models to white-noise data. As we did when using AR(1) and MA(1) processes as local approximations, we find evidence of the absence of local linear dependence in the sequence of log absolute returns.

As a conclusion, in a flexible modeling framework which nests time-varying linear dependence structure and time-varying mean and variance, the data choose as most appropriate a simple model with no linear dependence but with significant changes in the mean and in the variance of the time series. Figures 1 and 2 suggests that piecewise, on the intervals of homogeneity, the data are approximately a white noise.

D. *The Asymptotic Distribution of $T(n, X, \mathcal{M}_{\hat{\mu},\, \hat{\sigma}^2,\, \hat{f}_\psi})$*

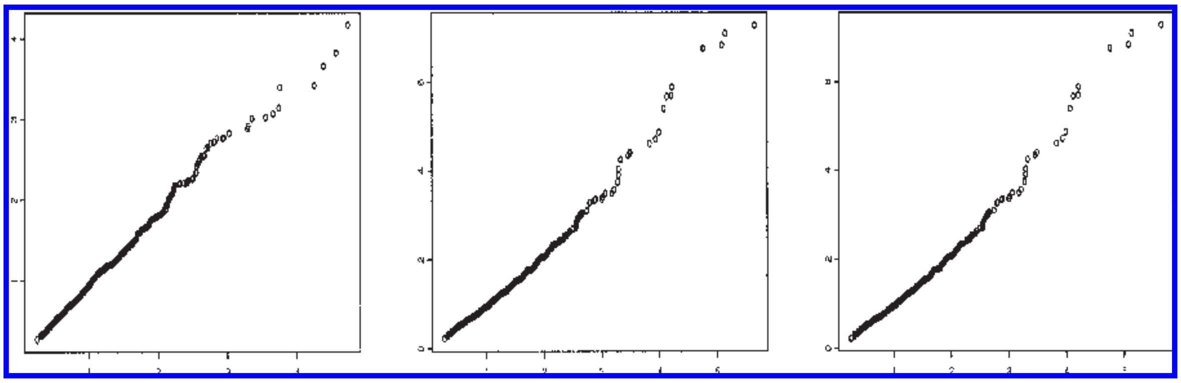
In constructing the homogeneity intervals that lead to the results presented in figure 2, we took the ARMA(1,1) linear model estimated in the initial part of an interval of homogeneity for the real data-generating process of the rest of the interval. In other words, we exchanged $T(n, X, \mathcal{M}_{\mu,\, \sigma^2, f_\psi})$ for $T(n, X, \mathcal{M}_{\hat{\mu},\, \hat{\sigma}^2 \hat{f}_\psi})$ but continued to use the asymptotic distribution of the first to decide on the extension of the homogeneity interval. The issue of the relationship between the distribution of the relevant test statistic $T(n, X, \mathcal{M}_{\hat{\mu},\, \hat{\sigma}^2,\, \hat{f}_\psi})$ and the theoretical distribution of $T(n, X, \mathcal{M}_{\mu,\, \sigma^2, f_\psi})$ has been addressed through a small simulation study. ARMA(1,1) linear models with parameters (μ, σ^2, f_ψ) similar to the ones obtained from the estimation procedure that produced the results in figure 2 were simulated. For every model 1000 samples were generated. The length of the sample was 250 (the same as in the previous analysis). For every sample, both $T(n, X, \mathcal{M}_{\mu,\, \sigma^2 f_\psi})$ and $T(n, X, \mathcal{M}_{\hat{\mu},\, \hat{\sigma}^2 \hat{f}_\psi})$ were calculated ($\hat{\mu}$, $\hat{\sigma}^2$, and $\hat{f}_\psi$ are the sample estimates of the true parameters μ, σ^2, and f_ψ). Figure 3 displays the simulation results for the models $\phi_1 = -0.8$, $\theta_1 = -0.8$ (left), $\phi_1 = 0.01$, $\theta_1 = 0.01$ (middle), and $\phi_1 = 0.8$, $\theta_1 = 0.8$ (right) ($\mu = -6$, $\sigma^2 = 1.4$ in all three cases). Each graph displays the QQ plot of the sample (of size 1000) of $T(n, X, \mathcal{M}_{\mu,\, \sigma^2,\, f_\psi})$ (on the x-axis) versus the sample (of the same length) of $T(n, X, \mathcal{M}_{\hat{\mu},\, \hat{\sigma}^2 \hat{f}_\psi})$ (on the y-axis). The graphs show that, for the purpose of testing at statistically common levels of confidence (say 95%), the distributions of the two statistics, $T(n, X, \mathcal{M}_{\mu, \sigma^2 f_\psi})$ and $T(n, X, \mathcal{M}_{\hat{\mu},\, \hat{\sigma}^2 \hat{f}_\psi})$ are practically identical.[19]

---

[18] The rest of the sample shows the same behavior.

[19] Removing the last 25 more extreme pairs in any of the graphs, that is, 2.5% of the sample, leaves us with an almost straight line, indicating very good correspondence of the quantiles (up to the 0.975th one) of the two distributions.

FIGURE 3.—QQ PLOT OF $T(n, X, \mathcal{M}_{\mu, \sigma^2, f_\psi})$ (ON THE x-AXIS) VERSUS $T(n, X, \mathcal{M}_{\hat\mu, \hat\sigma^2, \hat f_\psi})$ (ON THE y-AXIS) CORRESPONDING TO ARMA(1,1) MODELS $\phi_1 = -0.8$, $\theta_1 = -0.8$ (*LEFT*), $\phi_1 = 0.01$, $\theta_1 = 0.01$ (*MIDDLE*), AND $\phi_1 = 0.8$, $\theta_1 = 0.8$ (*Right*) ($\mu = -6$, $\sigma^2 = 1.4$ IN ALL THREE CASES)



$\hat\mu$, $\hat\sigma^2$, and $\hat f_\psi$ are the sample estimates of the true parameters μ, σ^2, and f_ψ. The length of the sample on which the test statistics were calculated was 250. 1000 samples of each model were generated.

V. Local Approximation by I.I.D. Processes

The results in the previous section suggest that a simple local approximation by processes free of second-order structure, characterized simply by the mean μ and variance σ^2, might be appropriate. In this section we present the results of an analysis that uses i.i.d. processes with changing unconditional mean and variance to locally describe the movements of the log absolute returns. We estimate the time-varying first and second unconditional moments, present some evidence supporting the choice of the simple local approximation under discussion, and briefly discuss its implications on describing the dynamics of the returns. Our analysis implies the following dynamics of the data suggested by the results in the previous section.

A. A Model for Returns

The results displayed in figures 1 and 2 indicate that our data $X_t = \log|r_t|$ can be modeled as independent[20] and suggest the following simple model:

$$X_t = \mu(t) + \sigma(t)\varepsilon_t, \tag{13}$$

where ε_t are i.i.d. with $E\varepsilon_t = 0$, $E\varepsilon_t^2 = 1$, and where the unconditional mean $\mu(t)$ and unconditional variance $\sigma^2(t)$ are functions of t. This model yields the following for the absolute returns and the returns:

$$|r_t| = h(t)^{1/2}\epsilon_t, \quad r_t = h(t)^{1/2}\epsilon_t S_t, \tag{14}$$

where $h(t) := e^{2\mu(t)}\, E(e^{2\sigma(t)\varepsilon_t})$ is the time-varying unconditional variance function, $\epsilon_t := e^{\sigma(t)\varepsilon_t}/[E(e^{2\sigma(t)\varepsilon_t})]^{1/2}$ are independent innovations with a time-dependent distribution, and $E\epsilon_t^2 = 1$. The sequence (S_t) is i.i.d., $S_t = -1, 1$ with probability 0.5.

In words, the returns are modeled as *independent random variables* with a time-varying unconditional variance[21] (if σ^2 is also changing through time, they might have other time-varying unconditional probabilistic characteristics). In particular, the returns are a nonstationary sequence of random variables. Note that if $\sigma(t)$ can be assumed close to constant, then the innovation sequence (ϵ_t) can be modeled as i.i.d., and the only time-varying feature of the returns is the unconditional variance.

To conduct the local analysis based on approximations with i.i.d. processes, we need to implement the steps described in section II. In particular, we need to make precise the test statistic to be used.

B. Test Statistic for I.I.D. Processes

Our earlier findings motivate a simpler test statistic than equation (6), namely,

$$\tilde{T}(n,X,\mu,\sigma^2) := \sup_{\lambda\in[0,\pi]} \left| \int_{-\pi}^{\lambda} \left(\frac{I_{n,X}(y)}{\sigma^2/2\pi} - \frac{\tilde\sigma^2}{\sigma^2} \right) dy \right|, \tag{15}$$

to help construct the homogeneity intervals. The asymptotic distribution which provides the critical values needed for hypothesis testing is given in corollary 3.2.

C. Estimation of the Time-Varying Unconditional Mean and Variance

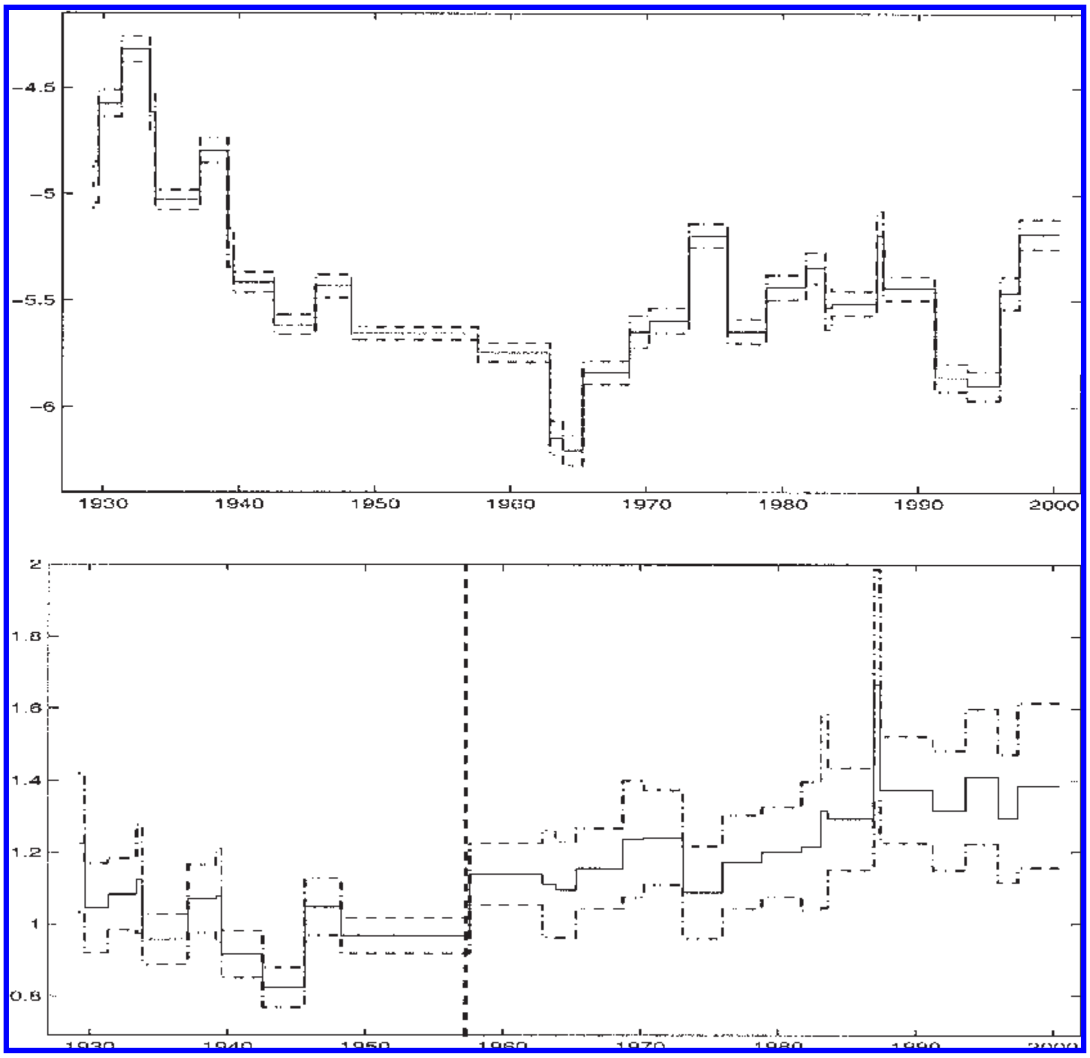
In approximating the movements of log absolute returns by i.i.d. processes, the dynamics of the data is concentrated in changes of the unconditional first two moments. Goodness-of-fit tests based on the statistic (ψ) were used as part of the methodology described in section II to produce homogeneity intervals on which the unconditional mean and variance were estimated.

---

[20] Further evidence supporting the choice of locally approximating the dynamics of the logarithm of the absolute return data by i.i.d. sequences is presented in section VI.

[21] The time-varying second moment is responsible for the structure present in the sample autocorrelation function (ACF) of absolute returns.

Figure 4.—The Estimated Time-Varying Unconditional Mean $\hat{\mu}$ (*Left*) and Variance $\hat{\sigma}^2$ (*Right*) of the Logarithm of Absolute Returns on S&P 500, X_t



The dotted lines are the 95% confidence intervals based on the CLT. The vertical line in the second graph marks the date when the definition of the S&P 500 index changed (see footnote 14).

Figure 4 displays the estimated unconditional mean $\hat{\mu}(t)$ and the estimated unconditional variance $\hat{\sigma}^2(t)$ of the logarithm of absolute values of daily returns. The confidence intervals are those given by the CLT applied to X_t and $[X_t - E(X_t)]^2$, respectively.

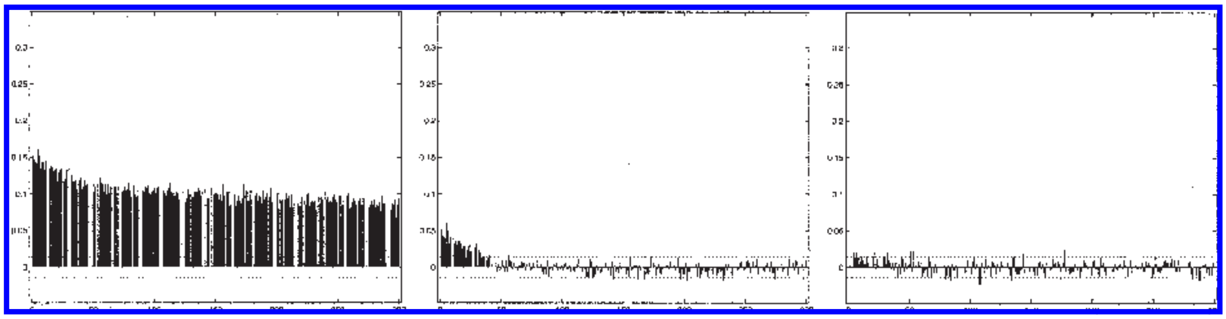
The top graph in figure 4 shows a very volatile decade between 1928 and 1938 (the high-mean period for log absolute returns ended rather abruptly around the beginning of the Second World War) followed by a mild downward trend until the middle of the 1960s (the postwar economic boom) and a general upward trend from then to the end of the sample. One can possibly see a certain connection between the higher levels of the mean of log absolute returns (hence higher unconditional variance of returns) and the 1973 oil crisis and the economic recessions in the beginning of the 1980s and 1990s. The strong market movements around the 1987 stock market crash are also visible. After a period of low mean in the middle of the 1990s, the end of the recent past period of economic expansion that spanned the 1990s is also characterized by a higher level of the mean of log absolute returns, that is, a higher unconditional variance of returns.

The bottom graph in figure 4 shows a significantly lower variance of the log absolute returns before the middle of 1950s, followed by a slight upward trend. The vertical line corresponds to the date when the definition of the S&P 500 index changed.

Figure 5 displays some evidence supporting the choice of i.i.d. sequences as local approximations of the dynamic of log absolute returns.

The first two graphs in figure 5 display the sample ACF for the logarithms of absolute values of daily returns before and after the data were centered at the mean $\hat{\mu}$ and standardized by the standard deviation $\hat{\sigma}$ estimated by our methodology. They show a strong reduction of the dependence present in the sample ACF (the residual dependence

FIGURE 5.—SAMPLE ACFS OF X_t, THE LOGARITHM OF THE ABSOLUTE RETURNS ON THE S&P 500, BEFORE (*LEFT*) AND AFTER (*MIDDLE*) SUBTRACTING THE ESTIMATED MEAN $\hat{\mu}$ AND STANDARDIZING WITH THE ESTIMATED STANDARD DEVIATION $\hat{\sigma}$; THE SAMPLE ACF OF THE ABSOLUTE VALUES OF CENTERED AND STANDARDIZED DATA, $X_t - \hat{\mu}(t)/\hat{\sigma}(t)$ (*RIGHT*)



is extremely small—less than 0.05—and confined to the first ≈30 lags).[22] The last graph of figure 5 presents the sample ACF of the absolute values of centered and standardized data, $|X_t - \hat{\mu}(t)|/\hat{\sigma}(t)$, and shows no linear dependence. The two sample ACFs corresponding to the estimated residuals (the middle and right graphs in figure 5); and showing almost no linear dependence suggest that *independent* sequences indeed provide good local approximations to the dynamics of the data.

D. The S&P 500 Volatility from 1928 to 2000

The intervals of homogeneity for the logarithms of absolute returns translate into intervals of homogeneity for the absolute returns.[23]

Figure 6 displays, in the graph at the top, the unconditional, time-varying annualized standard deviation the returns of the S&P 500 together with the returns themselves. In the two graphs at the bottom it displays the sample ACF for the absolute values of daily returns before and after the data were scaled by the standard deviation in the graph on top. The two sample ACF graphs show a strong reduction of the dependence. The remaining dependence is small, less than 0.15, and it is confined to the first ≈40 lags.

E. Intervals of Homogeneity Based on the Central Limit Theorem

We end this section with a discussion of the performance of our procedure on replacing the goodness-of-fit test based on the integrated periodogram with a very simple goodness of fit based on the CLT. The null hypothesis is that locally the data are independent with mean μ and variance σ^2. The test statistic is simply

$$\bar{T}(n,X,\mu,\sigma) := \frac{\bar{X} - \mu}{\sigma}. \tag{16}$$

The methodology outlined in section II and the test statistic (15) were used to produce the homogeneity intervals on which the first two unconditional moments were estimated. The estimation results are displayed in figure 7.

Though the tests based on the statistic $\bar{T}(n, X, \mu, \sigma)$ found more changes than the integrated periodogram, the overall pattern of change is the same. The overall amount of dependence in the absolute returns explained by the shifts in the first two unconditional moments (as measured by the residual correlation in sample ACF of the standardized time series) is practically the same.

As a conclusion, the integrated periodogram approach offers a simpler overall picture of the pattern of changes in the long time series and, more importantly, serves to motivate the assumption of locally independent log absolute returns on which the use of the test statistic $\bar{T}(n, X, \mu, \sigma)$ is based.

VI. A Simple Model for the Period 1957–2000

In this section a simple model for the log absolute returns and returns covering the period between 1957 and 2000 is discussed. A great deal of attention is devoted to checking the goodness of fit of the model to the data. The results confirm our choice of modeling the return data as locally i.i.d. sequences.

A. The Model for Returns Revisited

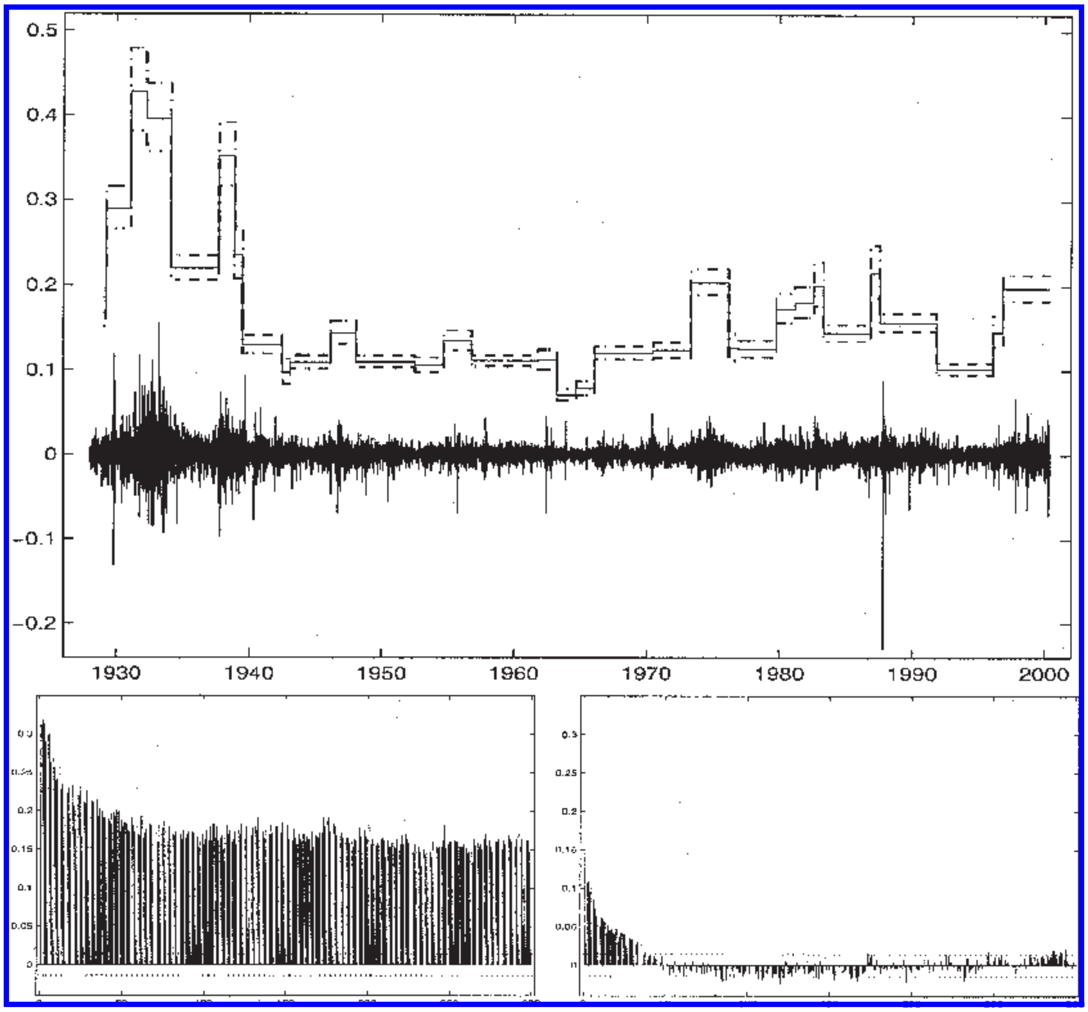
The bottom graph in figure 4 shows a change in the estimated variance of the log absolute returns occurring in the middle of the 1950s. The estimated change coincides in date with the change in the definition of the index. Before the mid 1950s the variance σ^2 of the log absolute returns was lower. Note also that after 1960 the value of σ^2 stayed roughly constant around 1.2 until the end of the 1980s, and has been around 1.4 since then. Moreover, the confidence intervals indicate that this increase may be not significant. This motivates the assumption of constant σ for the period

---

[22] A better approximation (than our rough step-function approximation) of the changing mean removes completely the linear dependence still present in the sample ACF of log absolute returns. For an analysis conducted on returns, see Drees and Stărică (2002) for the univariate case and Herzel, Stărică, and Tütüncü (2002) for the multivariate case.

[23] The methodology described in section II is applicable as long as the data have a finite fourth moment; see theorem 3.1. This condition is barely satisfied by the absolute returns, which have a negative tail index close to 4. Although strictly speaking theoretically feasible, our methodology does not produce meaningful results when applied directly to absolute returns.

FIGURE 6.—TOP: ESTIMATED TIME-VARYING UNCONDITIONAL STANDARD DEVIATION (ANNUALIZED) WITH 95% CONFIDENCE INTERVALS TOGETHER WITH THE RETURNS ON THE S&P 500; BOTTOM: SAMPLE ACFS OF THE ABSOLUTE VALUES OF RETURNS ON THE S&P 500 BEFORE (*LEFT*) AND AFTER (*RIGHT*) SCALING WITH THE STANDARD DEVIATION ABOVE



after 1957. We will see in the sequel that this assumption provides a good approximation to the data dynamic.

Hence, for the period between 1957 and the end of the sample, the model for log absolute returns (13) could be further simplified to

$$X_t = \mu(t) + \sigma\varepsilon_t, \tag{17}$$

where ε_t are i.i.d. with $E\varepsilon_t = 0$ and $E\varepsilon_t^2 = 1$. This yields the following model for the absolute returns and the returns:

$$|r_t| = h(t)^{1/2}\tilde{\varepsilon}_t, \quad r_t = h(t)^{1/2}\tilde{\varepsilon}_t S_t, \tag{18}$$

with $h(t) := e^{2\mu(t)}\, E(e^{2\sigma\, \varepsilon_t})$, $\tilde{\varepsilon}_t := e^{\sigma\, \varepsilon_t}/[E(e^{2\sigma\, \varepsilon_t})]^{1/2}$, $E\tilde{\varepsilon}_t^2 = 1$, $S_t = -1, 1$ with probability 0.5, and (ε_t) and (S_t) i.i.d. sequences.

In words, the returns are modeled as independent random variables with a time-varying unconditional variance. They form a nonstationary sequence, free of any dependence,[24] but with a marginal distribution that evolves through time. Moreover, the only changing probabilistic feature of the marginal distribution is the unconditional variance.

The reduction in the complexity of the model implies also a simplification of the statistical procedure for constructing the homogeneity intervals.

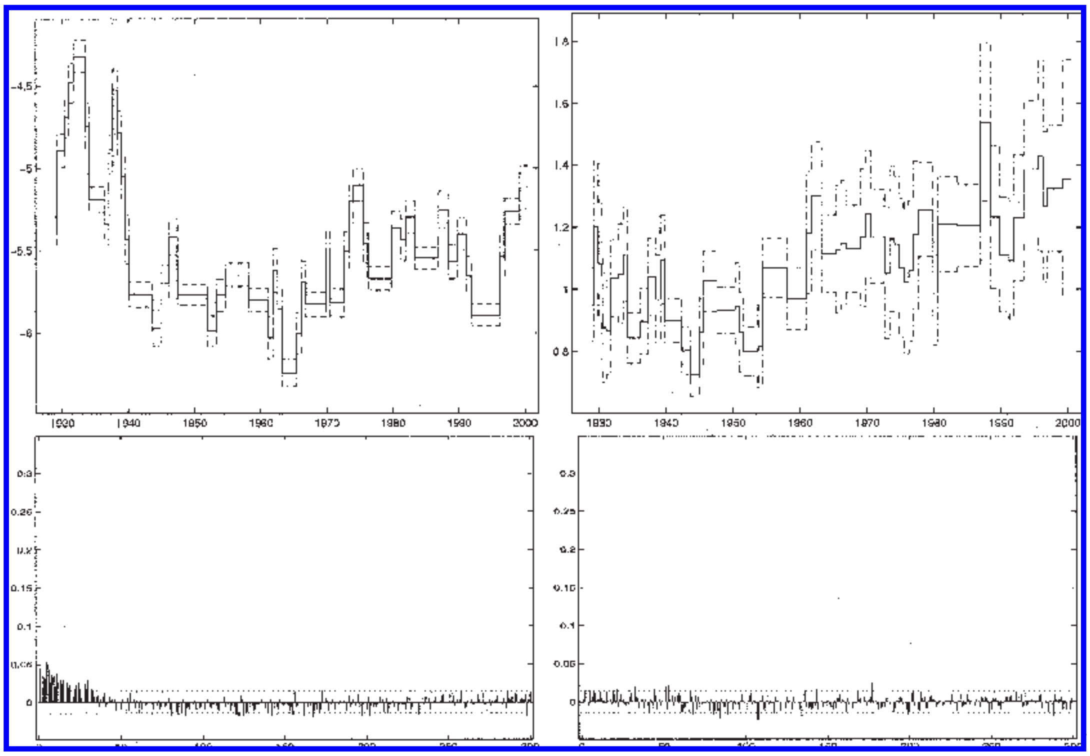
B. *The Test Statistic Revisited*

In the test statistic (15), the ratio $\tilde{\sigma}^2/\sigma^2$ is replaced by 1. In this way, only the changes in the mean can produce extreme values of the test statistic and hence cause a homogeneity interval to end. The result of estimating the

---

[24] Independent nonstationary sequences can display significant sample ACF. In particular, the long-memory effect in volatility can occur for independent sequences with a time-varying unconditional variance. For more details on this issue, see Mikosch and Stărică (2004).

FIGURE 7.—TOP: THE ESTIMATED TIME-VARYING UNCONDITIONAL MEAN $\hat{\mu}(t)$ (*LEFT*) AND VARIANCE $\sigma^2(t)$ (*RIGHT*) OF THE LOGARITHM OF THE ABSOLUTE RETURNS ON THE S&P 500; BOTTOM: SAMPLE ACFS OF THE ESTIMATED RESIDUALS $[X_t - \hat{\mu}(t)]/\hat{\sigma}(t)$ (*LEFT*) AND THEIR ABSOLUTE VALUES (*RIGHT*)



The intervals of homogeneity were constructed using the test statistic $\bar{T}(n, X, \mu, \sigma)$ defined in equation (16). The dotted lines are the 95% confidence intervals based on the CLT.

function $\mu(t)$ is practically identical to the one in figure 4, and hence we do not redisplay it.

In the sequel we evaluate carefully how well the model (17) fits the log absolute returns between 1957 and 2000.[25] The assessment of the goodness of fit will also yield a validation of the choice of modeling the return data as locally i.i.d.

C. Goodness-of-Fit Analysis

The goodness-of-fit analysis is based on the sequence of *estimated residuals*

$$\hat{\varepsilon}_t := X_t - \hat{\mu}(t)\,, \qquad (19)$$

where the time-varying first unconditional moment $\mu(t)$ is estimated on the homogeneity intervals constructed using the test statistic described above.

We begin with an evaluation of the assumption that ε_t are independent. Toward this end we assess the dependence in the sequences $\hat{\varepsilon}_t$ and $|\hat{\psi}_t|$. The linear dependence is measured via the sample ACF, and the nonlinear dependence is evaluated by means of copulas, a notion to be described shortly.

Independent Estimated Residuals: The top left graph in figure 8 displays a plot of the estimated residuals. Visual inspection shows no signs of dependence. The apparent pattern in the lower part of the graph is due to the discrete nature of the prices (the minimum price increment was big in the beginning of the period compared to its end). The top right graph displays the histogram of the estimated returns, showing a distribution skewed to the right. The bottom left graph is the sample ACF of the logarithm of absolute returns between 1957 and 2000. It displays the so-called long-memory effect in volatility. The last two graphs are the sample ACFs of the sequence (19) and their absolute values.

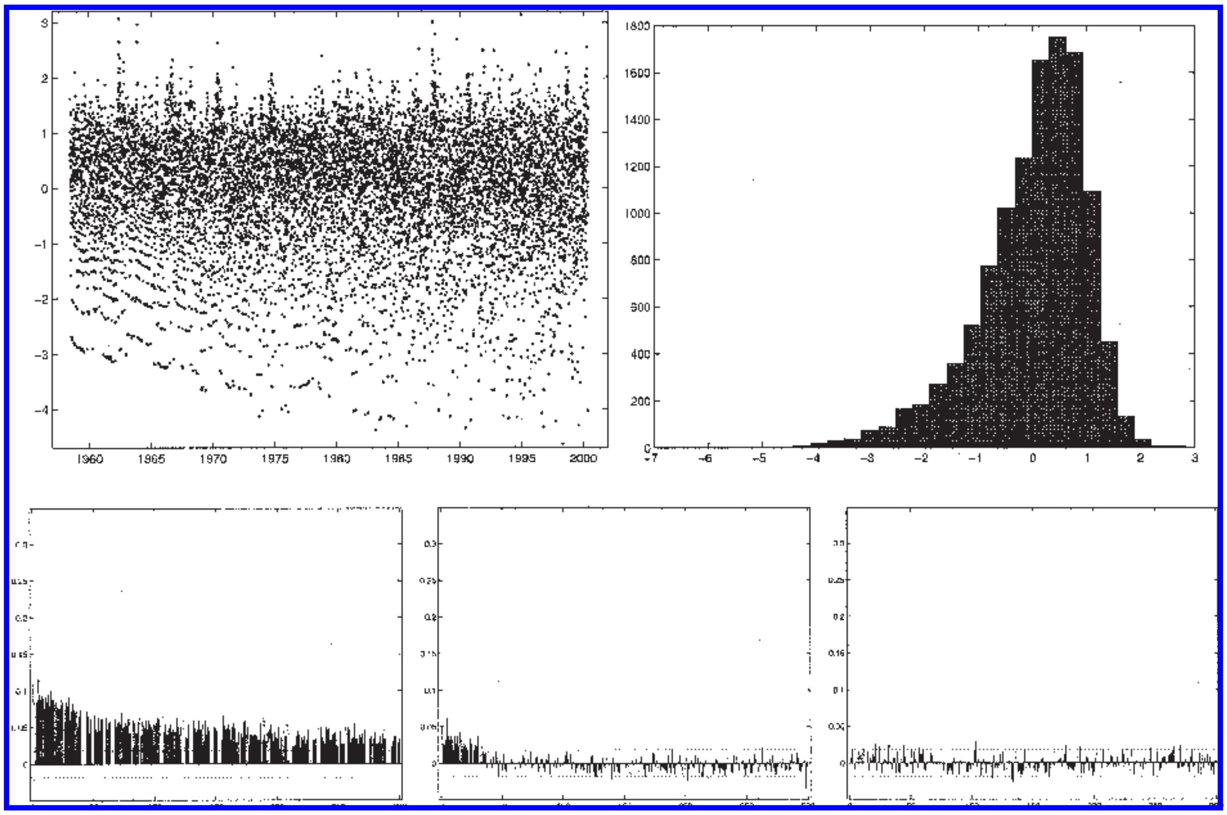
The sample ACFs of the estimated residuals $\hat{\varepsilon}_t$ as well as their absolute values $|\hat{\varepsilon}_t|$, shown in the bottom row of figure 8 (middle and right), are close to being statistically insignificant, indicating that almost no linear dependence remains in the time series of residuals and in their absolute values.

To search for possible patterns of nonlinear dependence, it is most useful to have a look at copulas of estimated residuals paired with lagged estimated residuals $(\hat{\varepsilon}_t, \hat{\varepsilon}_{t+i})$ and the pairs of absolute values $(|\hat{\varepsilon}_t|, |\hat{\varepsilon}_{t+i}|)$, $i = 1, 2, \ldots$. Before showing the results of this assessment, let us say a few words about the notion of copula [for more details see Nelsen (1999)].

---

[25] Moving the beginning of the subsample to 1960, as the estimated variance in figure 7 might suggest, does not change the nature of the results in any way. Hence we preferred the longer subsample.

Figure 8.—Top: Plot (*Left*) and Histogram (*Right*) of the Estimated Residuals $\hat{\varepsilon}_t$ Based on the Model (8) fitted to the period 1957–2000; Bottom: Sample ACFs of X_t, the Logarithm of the Absolute Returns of the S&P 500, Period 1957–2000 (*Left*); Sample ACF for the Estimated Residuals $\hat{\varepsilon}_t$ (*Middle*); Sample ACF of the Absolute Values of Residuals $|\hat{\varepsilon}_t|$ (*Right*)



Middle: Due to the rough approximation of the mean of the time series by a step function, the first ≈25 lags are slightly significant. Right: The graph shows no properties of the variance of the residuals.

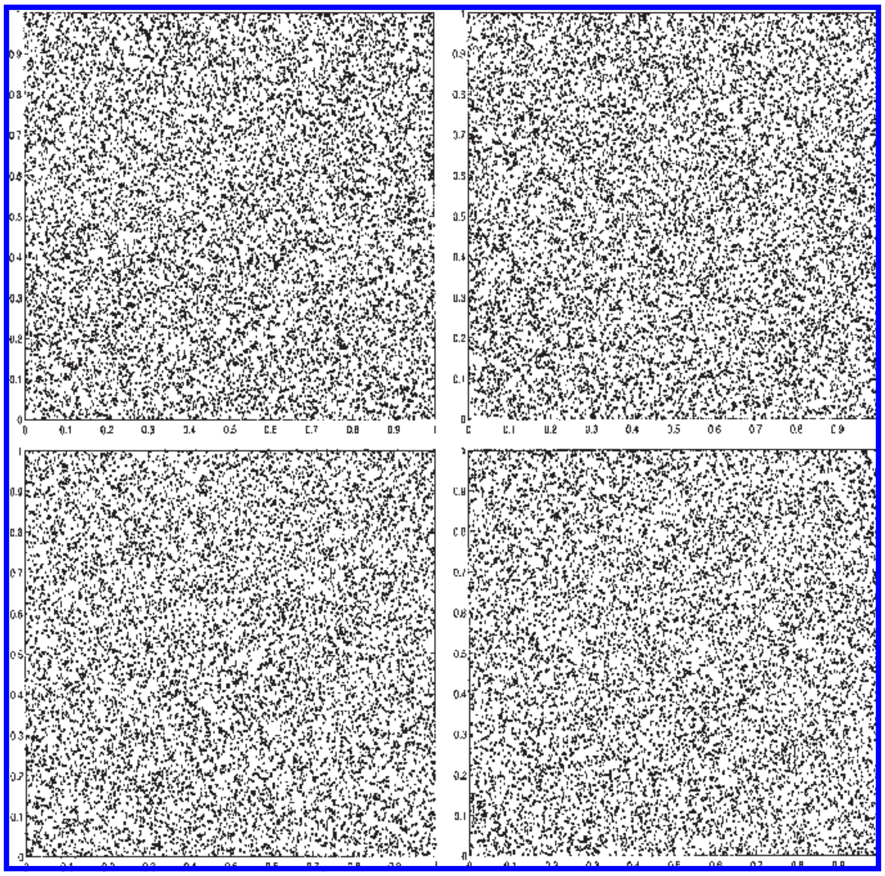
The joint distribution of a pair of random variables (U, V) is uniquely determined by the marginal distribution of the coordinates F_U and F_V and by their copula, that is, the distribution on the unit square of ($F_U(U)$, $F_V(V)$). Hence, it is the copula that provides the complete description of the dependence structure between the marginal random variables. Moreover, U and V are independent if and only if their copula is the uniform copula. Graphically, this corresponds to a uniform filling of the unit square by the pairs ($F_U(U)$, $F_V(V)$). Hence, a simple but very informative way of assessing the independence of the coordinates of a bivariate random vector is looking at realizations of its copula. The appearance of a uniformly covered unit square supports the assumption of independence the presence of patterns indicates dependence.

To obtain the copula associated with the mentioned bivariate random vectors, we first transformed the residuals and their absolute values into uniform random variables using the empirical distribution functions $\hat{F}_{\hat{\varepsilon}}$ and $\hat{F}_{|\hat{\varepsilon}|}$. Then we produced the scatterplots (Y_t, Y_{t+i}) for $Y_t = \hat{F}_{\hat{\varepsilon}}(\hat{\varepsilon}_t)$ (figure 9, left) and $Y_t = \hat{F}_{|\hat{\varepsilon}|}(|\hat{\varepsilon}_t|)$ (figure 9, right), and for $i = 1$ (figure 9, top) and $i = 2$ (figure 9, bottom). As mentioned, a uniform filling of the unit square is interpreted as evidence of *independent* components. The graphs in figure 9 reveal a uniform covering of the unit square for transformed residuals paired with their first and second lagged values. The same behavior is apparent for higher lags *i.* This finding, together with the previous evidence on the linear dependence in the time series of estimated residuals, confirms that the assumption of *independent innovations* provides a reasonable approximation for the dynamics of the log absolute data under scrutiny.

Identically Distributed Residuals: To check the relevance of the assumption that the innovations are identically distributed, we divided the sample of residuals into three subsamples of equal length, the first corresponding roughly to the period 1958–1972, the second to 1972–1986, and the third to 1986–2000. Due to the discrete nature of the observations, only the residuals bigger than -2 were included in the subsamples. Figure 10 displays the QQ plots corresponding to the three pairs of subsamples. Note that we are comparing, on the three subsamples, the conditional distribution of the residual sequence given that a value bigger than -2 was taken. The three graphs support the hypothesis of identically distributed innovations of the model (17).

To summarize, the simple model (17) that assumes the returns to be independent with a time-varying unconditional variance describes well the data from the period 1957–2000.

FIGURE 9.—SCATTERPLOTS $(Y_t,\ Y_{t+1})$ (*TOP*) AND $(Y_t,\ Y_{t+2})$ (*BOTTOM*) FOR $Y_t = \hat{F}_{\hat{\varepsilon}}(\hat{\varepsilon}_t)$ (*LEFT*) AND $Y_t = \hat{F}_{|\hat{\varepsilon}|}(|\hat{\varepsilon}_T|)$ (*RIGHT*)



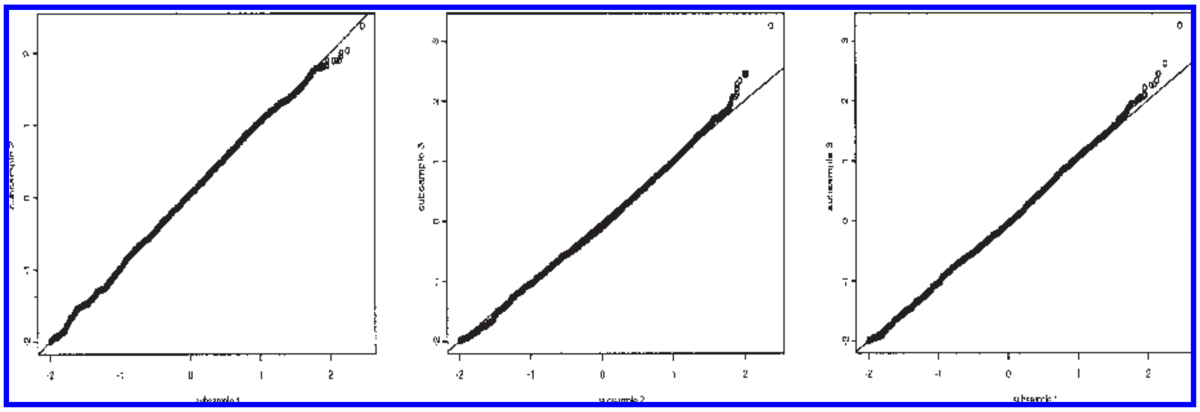
The empirical distribution function estimated on the sample $\hat{F}_{\hat{\varepsilon}}$ ($\hat{f}_{|\hat{\varepsilon}|}$) was used to transform the marginal distribution of the data to uniform. A uniform filling of the unit square is interpreted as evidence of independent components.

Due to the rather coarse estimation method which approximates smooth functions with piecewise constant ones, the residuals present slight traces of linear dependence. The absolute values of the residuals seem independent. Practically speaking, all dynamics of the log absolute return (the return) time series seem to be concentrated in shifts of the unconditional mean (the variance).

VII. Forecasting Comparison: Nonstationary versus Stationary and Long Memory

In the stationary framework, a sample ACF behavior like that shown in the left graphs in figure 5 and figure 6 will be interpreted as evidence of long memory. Hence we are facing a modeling choice for $X_t = \log|r_t|$. The choice is

FIGURE 10.—PAIRWISE QQ PLOTS OF THREE SUBSAMPLES OF EQUAL LENGTH, ROUGHLY CORRESPONDING (LEFT TO RIGHT) TO THE PERIODS 1958–1972, 1972–1986, AND 1986–2000



Due to the discrete nature of the data, only the returns bigger than −2 were considered.

between a stationary long-memory model and a nonstationary model with the dynamics mainly concentrated in changes of the mean. One possible way of solving this dilemma is to compare the forecasting behavior of two paradigms on the data at hand. Because our approach is to describe the volatility directly by analyzing the sequence of absolute returns, a natural choice for a long-memory stationary model is the fractionally ARIMA class introduced by Granger and Joyeux (1980) and Hosking (1981).

The process (X_t) is said to be a FARIMA (p, d, q) with $d \in (0, 0.5)$ if (X_t) is stationary and satisfies the difference equation

$$\phi(B)\nabla^d X_t = \theta(B) Z_t, \tag{20}$$

where (Z_t) is white noise and ϕ, θ are polynomials of degree p, q respectively. The operator ∇^d is defined by

$$\nabla^d := (1 - B)^d = \sum_{j=0}^{\infty} \pi_j B^j, \tag{21}$$

where

$$\pi_j = \frac{\Gamma(j - d)}{\Gamma(j + 1)\Gamma(1 - d)} = \prod_{0<k\leq j} \frac{k - 1 - d}{k}, \quad j = 1,2, \ldots .$$

The data used in the out-of-sample comparison are the logarithms of the absolute values of daily returns in the interval 1961–2000 (the data from 1957 to 1960 were used for initial estimation of the models). A FARIMA(1,d,1) model (LM) was estimated on the first 1000 observations (corresponding roughly to the period 1957–1960) and reestimated every month (that is, every 20 observations) using all the past observations.

A. Comparison of Forecasts of Daily Log Absolute Returns

With the estimated long-memory model, predictions (f^{LM}) for the future values of $X_t = \log|r_t|$ were made every month (that is, every 20 observations). The maximal forecasting horizon p was 200 days ahead. The other model used, which will be referred as the shifts-in-the-mean model (SM), is described by equation (17). The observations anterior to the date when a forecast was made were used for determining the (then) current interval of homogeneity. The forecasts for the future values of X_t, (f^{SM}), were simply the estimated means on this homogeneity interval and hence do not change with the horizon.

One way of comparing the two forecasts would be by assessing the orthogonality of one forecast error (at horizon p) to the other forecast. Concretely, one can test something less general, namely, whether one forecast error is uncorrelated with the other forecast. This can be accomplished by means of a regression. For example, to test if the SM forecasts and the LM forecast errors are uncorrelated, one would test whether $\alpha = 0$, $\beta_1 = 0$ in the regression

$$X_{t+p} - f_{t+p}^{\mathrm{LM}} = \alpha + \beta_1 f_{t+p}^{\mathrm{SM}} + \varepsilon_t . \tag{22}$$

However, given the possibly nonstationary nature of the time series, the assumption of ergodic stationarity of the regressors and dependent variables (needed for the well-functioning of the GMM machinery) is likely to be violated (f_{t+p}^{SM} is close to a piecewise constant function).

To address this possible problem we reformulate our test. Testing whether $\alpha = 0$, $\beta_1 = 0$ in equation (22) is equivalent to testing whether

$$\alpha = 0, \quad \beta_1 = 0, \quad \beta_2 = 1 \tag{23}$$

in the following regression:

$$X_{t+p} - X_t = \alpha + \beta_1(f_{t+p}^{\mathrm{SM}} - X_t) + \beta_2(f_{t+p}^{\mathrm{LM}} - X_t) + \varepsilon_t . \tag{24}$$

Notice that testing for

$$\alpha = 0, \quad \beta_1 = 1, \quad \beta_2 = 0 \tag{25}$$

in the same regression would be equivalent to verifying that the SM forecast error $X_{t+p} - f_{t+p}^{\mathrm{SM}}$ is uncorrelated with the FARIMA forecast f_{t+p}^{LM}. For this regression the violations of the assumption of ergodic stationarity of the regressors and dependent variables are likely to be less severe. Indeed, the vector $(X_{t+p} - X_t, f_{t+p}^{\mathrm{SM}} - X_t, f_{t+p}^{\mathrm{LM}} - X_t)$ is stationary and ergodic on every interval on which the volatility process $\sigma(t)$ is constant, that is, on any interval of homogeneity. Because under the null hypothesis (25) the error term ψ_t in the regression (24) equals the forecast error $X_{t+p} - f_{t+p}^{\mathrm{SM}}$—which is orthogonal to anything known at date t, including $f_t^{(i)} - X_t$, $i = 1, 2$—the regressors are guaranteed to be orthogonal to the error term [a similar statement holds under the null hypothesis (23)].

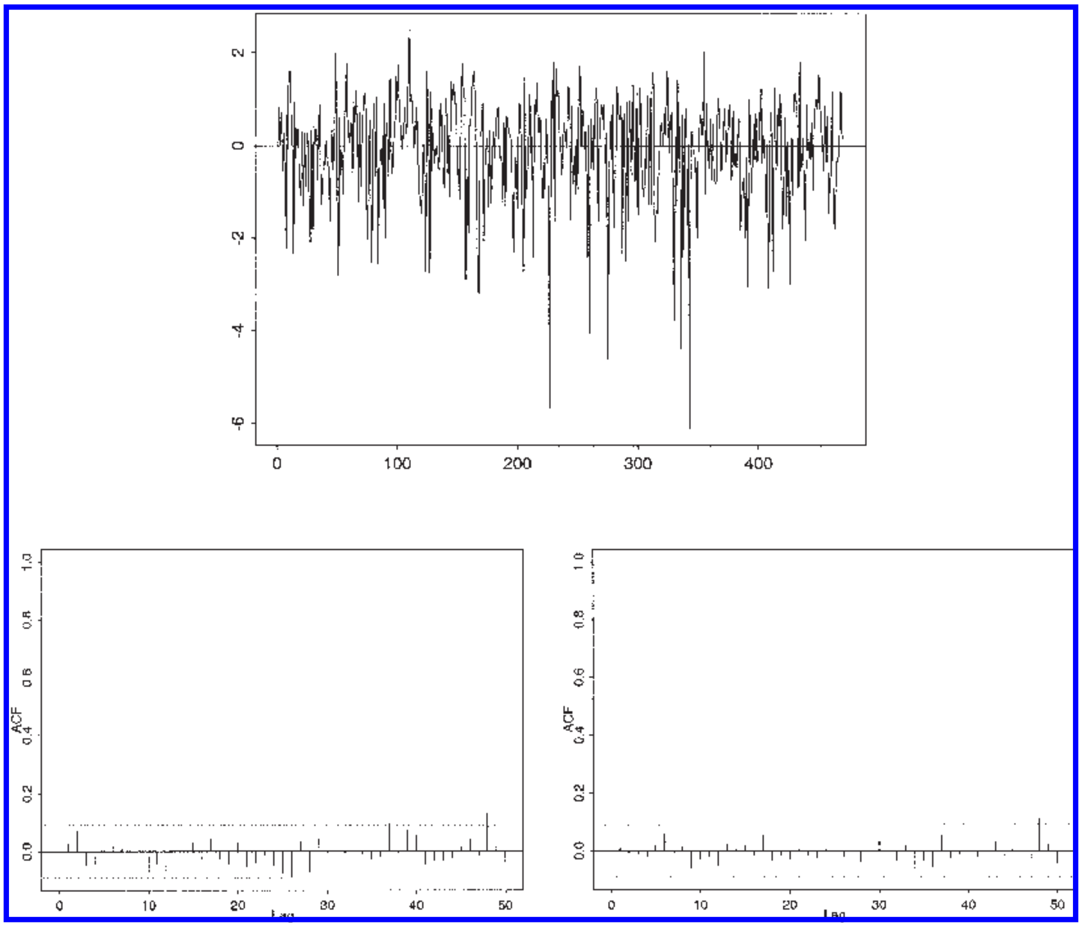
Even more, figure 11 supports the hypothesis of uncorrelated forecast errors (the forecast errors $X_{t+p} - f_{t+p}^{\mathrm{LM}}$ have a similar behavior), and hence it appears that an OLS estimate would suffice. Note that the regression (24) is closely related to the so-called *forecast encompassing equation*

$$X_{t+p} = \alpha + \beta_1 f_{t+p}^{\mathrm{SM}} + \beta_2 f_{t+p}^{\mathrm{LM}} + \varepsilon_t , \tag{26}$$

which cannot be employed, due to the possible nonstationarity both of the regressors and of the dependent variables.

The p-values of the F-test Wald statistic corresponding to $H_0 : \alpha = 0$, $\beta_1 = 0$, and $\beta_2 = 1$ and to $H_0 : \alpha = 0$, $\beta_1 = 1$, and $\beta_2 = 0$, respectively, are reported in table 1. For most of the forecast horizons the hypothesis of orthogonality of the SM forecast to the LM forecast errors is rejected, while the hypothesis of orthogonality of the LM forecast to the SM forecast errors remains unchallenged.

Figure 11.—Top: The Forecast Errors $X_{t+p} - f^{\mathrm{SM}}_{t+p}$ Based on the Model (17) Corresponding to the Period 1961–Present (the Period 1957–1960 Is Used for the Preliminary Estimation of the LM Model); Bottom: (*Left*) Sample ACF for the Forecast Errors $X_{t+p} - F^{\mathrm{SM}}_{t+p}$; (*Right*) Sample ACF of Absolute Forecast Errors $|X_{t+p} - f^{\mathrm{SM}}_{t+p}|$.



The graphs suggest that the forecast errors are uncorrelated and homoskedastic.

A possible critique of the previous analysis could be that we have used daily data as a measure against which to compare the performance of the two methodologies. As figure 8 shows, the daily data contain a large amount of idiosyncratic noise added to the signal of interest for us, the mean. In other words, daily data might not provide such a good check, as they are a poor measure of the mean.

Table 1.—Comparison of Forecasting Performance between the LM Model and SM Model

	p-Value of the Wald Statistic for			*p*-Value of the Wald Statistic for	
Horizon *p* (days)	$H_0 : \alpha = 0, \beta_1 = 0, \beta_2 = 1$, i.e., $f^{\mathrm{SM}}_{t,p} \perp e^{\mathrm{LM}}_{t,p}$	$H_0 : \alpha = 0, \beta_1 = 1, \beta_2 = 0$, i.e., $f^{\mathrm{LM}}_{t,p} \perp e^{\mathrm{SM}}_{t,p}$	Horizon *p* (days)	$H_0 : \alpha, \beta_1 = 0, \beta_1 = 0, \beta_2 = 1$, i.e., $f^{\mathrm{SM}}_{t,p} \perp e^{\mathrm{LM}}_{t,p}$	$H_0 : \alpha = 0, \beta_1 = 1, \beta_2 = 0$, i.e., $f^{\mathrm{LM}}_{t,p} \perp e^{\mathrm{SM}}_{t,p}$
10	0.00	0.21	110	0.03	0.22
20	0.03	0.27	120	0.06	0.27
30	0.00	0.31	130	0.00	0.06
40	0.05	0.67	140	0.06	0.40
50	0.01	0.49	150	0.02	0.27
60	0.09	0.23	160	0.03	0.08
70	0.01	0.32	170	0.00	0.48
80	0.09	0.39	180	0.06	0.16
90	0.01	0.44	190	0.00	0.59
100	0.01	0.12	200	0.06	0.23

The LM process is reestimated every 20 days using all past observations. The Wald statistic of the *F*-ratio test is calculated under the two alternatives, and the *p*-values are reported. A small *p*-value is a signal of the failure of the null. Overall the table shows a better performance of the SM model in forecasting. $e^{\mathrm{LM}}_{t,p} : = X_{t+p} - f^{\mathrm{LM}}_{t+p}$ (the forecast error of the long memory model at horizon *p*), and $e^{\mathrm{SM}}_{t,p} := X_{t+p} - f^{\mathrm{SM}}_{t+p}$ (the forecast error of the SM model at horizon *p*).

TABLE 2.—COMPARISON OF FORECASTING PERFORMANCE BETWEEN THE LM MODEL AND SM MODEL

Horizon p (days)	$\mathrm{MSE}^{\mathrm{SM}}(p)/\mathrm{MSE}^{\mathrm{LM}}(p)$		Horizon p (days)	$\mathrm{MSE}^{\mathrm{SM}}(p)/\mathrm{MSE}^{\mathrm{LM}}(p)$	
	S&P 500	NASDAQ		S&P 500	NASDAQ
10	0.99	1.02	110	0.91	0.56
20	0.99	0.88	120	0.91	0.54
30	0.98	0.82	130	0.91	0.53
40	0.96	0.75	140	0.91	0.52
50	0.95	0.70	150	0.91	0.51
60	0.93	0.65	160	0.91	0.51
70	0.93	0.63	170	0.91	0.51
80	0.92	0.61	180	0.91	0.51
90	0.91	0.59	190	0.91	0.51
100	0.91	0.57	200	0.91	0.51

The LM process is reestimated every 20 days using all the previous observations in the sample. The ratio $\mathrm{MSE}^{\mathrm{SM}}(p)/\mathrm{MSE}^{\mathrm{LM}}(p)$ is reported. A ratio smaller than 1 at horizon p indicates that the volatility forecast of the SM model at horizon p is more precise than that of the LM model. The figure shows overall better longer-horizon volatility forecast performance of the SM model.

B. *Comparison of Forecasts of the Log Absolute Returns over Various Time Intervals*

To answer this possible critique and in order to get a more complete picture, aggregated data at various horizons p were also used as an alternative control measure. Define

$$X_{t,p} = \sum_{i=1}^{p} X_{t+i}, \quad \bar{f}^{\,\mathrm{SM}}_{t,p} = \sum_{i=1}^{p} f^{\mathrm{SM}}_{t+i}, \quad \bar{f}^{\,\mathrm{LM}}_{t,p} = \sum_{i=1}^{p} f^{\mathrm{LM}}_{t+i}. \tag{27}$$

Note that $f_{t,p}$ are forecasts of the mean of $X_{t,p}$. For $p > 1$, through averaging, some of the idiosyncratic noise in the daily data is canceled, yielding a better measure against which to check the quality of the two forecasts.

We calculated and compared the MSE of the two methods, defined as

$$\mathrm{MSE}^{*}(p) := \sum_{t=1}^{n} (X_{t,p} - \bar{f}^{\,*}_{t,p})^2 \tag{28}$$

with * standing for SM or LM.

The results are presented in table 2, where the ratio[26]

$$\mathrm{MSE}^{\mathrm{SM}}(p)/\mathrm{MSE}^{\mathrm{LM}}(p)$$

is given for $p = 10, 20, \ldots, 200$. Besides the S&P 500 data, the MSE analysis has been performed also on the returns of the NASDAQ index between October 12, 1984 and November 8, 2002. The results for both series seem to support the conclusion that the SM model outperforms the LM model in forecasting.

VIII. Forecasting Comparison: Unconditional and Nonstationary versus Conditional and Stationary

As this paper addresses the issue of volatility and proposes a novel modeling paradigm, a comparison with the Garch framework, the current market leader in volatility modeling, is inevitable. In this section, we present the results of a forecasting comparison between our model (18) and a Garch (1,1) model.[27]

In the stationary, conditional framework, the working assumption is that a stationary Garch(1,1) process

$$r_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \tag{29}$$

where the innovations (ε_t) are i.i.d., mean 0, variance 1 (Student-*t* distributed in the case of the Student-*t* Garch), is a good approximation of the data-generating process. In the nonstationary, unconditional setup, we assume that the model (6.2) provides a good description of the daily returns.

We begin with an evaluation of the relevance of the assumption of a stationary Student-*t* Garch(1,1) data-generating process (DGP) for the S&P 500 return data. The relevance of a general Garch(1,1) process, without a concrete specification of the conditional distribution, as DGP for the S&P 500 return series is discussed in detail in Stărică (2003).[28]

A. *Student-t Garch(1,1) Model as Data-Generating Process for Returns*

The evaluation is based on the time evolution of the estimated parameters of the model. More concretely, a Student-*t* Garch(1,1) process was reestimated periodically (every 100 days), using maximum likelihood, in three different setups. In the first setup, the estimation was done on a sample which included 1000 past observations, in the second setup the sample contained 2000 past returns, and in the third setup, all past observations from the beginning of the sample were used.

Figure 12 displays the results of the estimation. The estimated parameters are plotted together with the 95%

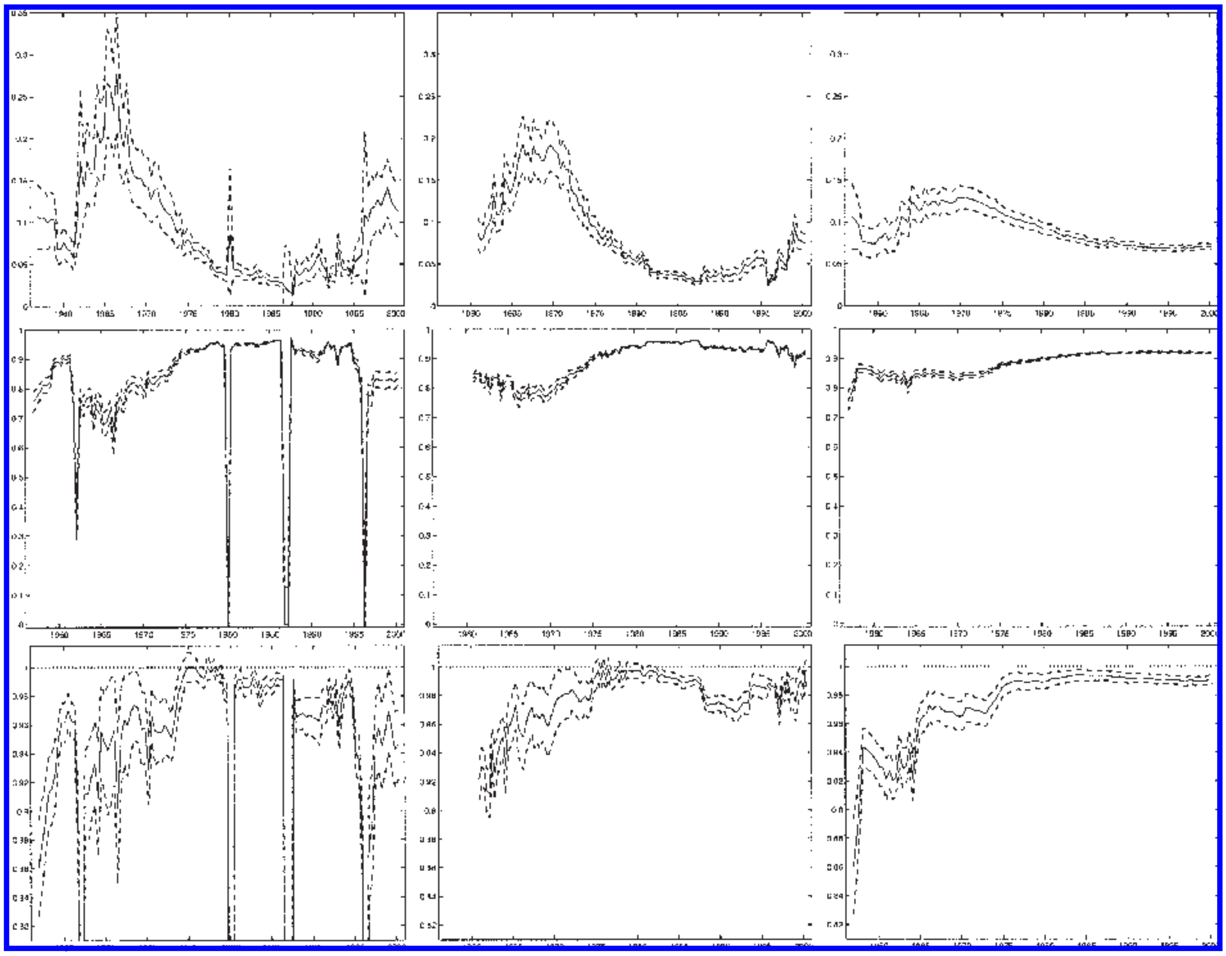
---

[26] Although the ratios in table 2 are sensibly different from 1 (especially in the case of the NASDAQ index), their statistical significance is difficult to assess. Tests of statistical significance for the differences of MSEs have been developed in the literature [see Diebold and Mariano (1995), West (1996), and Harvey, Leybourne, and Newbold (1997) among others]. The null hypothesis is that $Ed_{t,p} = 0$, where $d_{t,p} : = \mathrm{SFE}^{\mathrm{LM}}_{t,p} - \mathrm{SFE}^{\mathrm{SM}}_{t,p}$ is the loss-differential series, $\mathrm{SFE}^{\mathrm{LM}}_{t,p} := (X_{t,p} - \bar{f}^{\,\mathrm{LM}}_{t,p})^2$, and $\mathrm{SFE}^{\mathrm{SM}}_{t,p} := (X_{t,p} - \bar{f}^{\,\mathrm{SM}}_{t,p})^2$. However, to the best of our knowledge, all tests assume stationarity and short memory of the loss-differential series. Neither of these assumptions holds in the case at hand. Recall that we are trying to distinguish between a null of long memory and an alternative of nonstationarity (shifts in the mean).

[27] Both a Student-*t* Garch, estimated with an exact maximum likelihood approach, and a general Garch(1,1) model with a nonspecified conditional distribution, estimated with a quasi maximum likelihood approach, were used. With this comparison we aim at shedding some light on the the relevance of the nonstationary, unconditional approach as compared with the stationary, conditional modeling paradigm. We chose to specifically include the Student-*t* Garch(1,1) model, following the wide consensus that seems to exist in the financial econometric literature that this model provides an adequate description of return data.

[28] There both the analysis of the estimated model parameters and the forecasting performance clearly show the inadequacy of the Garch(1,1) model.

FIGURE 12.—THE ESTIMATED COEFFICIENTS α_1 (TOP), β_1 (MIDDLE), AND $\alpha_1 + \beta_1$ (BOTTOM) EQUATION (29), TOGETHER WITH 95% CONFIDENCE INTERVALS



The model was reestimated every 100 days on a sample containing 1000 past observations (left column), 2000 past observations (middle column), or all the past observations (right column). The graphs reject the hypothesis of a stationary Student-*t* Garch(1,1) data-generating process.

confidence intervals from the maximum likelihood estimation procedure.[29]

If the data-generating process for the S&P 500 return sample under scrutiny is a Student-*t* Garch(1,1) model, the parameters estimated on a moving window of constant length (figure 12, all rows, left and middle columns) should not vary significantly through time. Even more, the estimated values obtained using the whole sample (figure 12, all rows, right column) should not differ significantly from the ones based only on a part of the sample.

The plots in the bottom row of figure 12 show the so-called IGARCH effect as defined and discussed in Mikosch and Stărică (2004), that is, the sum $\alpha_1 + \beta_1$ stays away from 1 when estimated on shorter samples but approaches 1 when the sample size grows. In that paper the authors argue that a possible reason for this behavior is the nonstationarity of longer samples.

The plots in figure 12 show clear signs of instability of the model parameters, rejecting the assumption of a stationary Student-*t* Garch(1,1) data-generating process. This finding is consistent with the results in Stărică (2003), where a Garch(1,1) is fitted to the S&P 500 returns series between 1957 and 2003 using the quasi maximum likelihood. In contrast with the estimation yielding the parameters in figure 12, the quasi maximum likelihood assumes, possibly erroneously, a normal conditional distribution and allows for the possible misspecification by using corrected, larger confidence intervals (see Straumann & Mikosch, 2005). The estimated quasi maximum-likelihood parameters also display a clear time evolution incompatible with the assumption of a Garch(1,1) data-generating process.

B. *Comparison of Longer-Horizon Volatility Forecasts*

Finally, we compare the volatility forecasting results of the two methodologies. As in the previous comparison of MSEs, aggregated data at various horizons p are used as a more reliable measure against which the performances are evaluated.

---

[29] Note that the exact ML confidence intervals are much narrower than the quasi-likelihood confidence intervals which allow for possible misspecification of the conditional distribution. See Stărică (2003).

In more detail, assuming a Garch(1,1) data-generating process (29) that also satisfies $\alpha_1 + \beta_1 < 1$,[30] it follows that the minimum mean squared error (MSE) forecast for Er^2_{t+p}, the square daily return corresponding to the pth day in the future

$$\sigma^{2,\mathrm{Garch}}_{t+p} := E_t r^2_{t+p} = \sigma^2_{\mathrm{Garch}} + (\alpha_1 + \beta_1)^{p-1}(\sigma^2_t - \sigma^2_{\mathrm{Garch}}) , \tag{30}$$

where $\sigma^2_{\mathrm{Garch}} := \alpha_0/(1-\alpha_1 - \beta_1)$ is the unconditional variance. The minimum-MSE forecast for $(r_{t+1} + \cdots + r_{t+p})^2$, the square of the return over the next p days, is then given by

$$\bar{\sigma}^{2,\ \mathrm{Garch}}_{t,p} := \mathrm{E}_{\mathrm{t}}(\mathrm{r}_{\mathrm{t+1}} + \cdots + \mathrm{r}_{\mathrm{t+p}})^2 = \sigma^{2,\ \mathrm{Garch}}_{\mathrm{t+1}} + \cdots + \sigma^{2,\ \mathrm{Garch}}_{t+p} . \tag{31}$$

Under the model (18), the forecast for Er^2_{t+p} is given by

$$\sigma^{2,\mathrm{SM}}_{t+p} := Er^2_t = h(t) , \tag{32}$$

with $h(t)$ as defined by (18), whereas the forecast for $E(r_{t+1} + \cdots + r_{t+p})^2$, the variance of the next p aggregated returns, is simply

$$\bar{\sigma}^{2,\mathrm{SM}}_{t,p} := pEr^2_t . \tag{33}$$

Define the following measure of the realized volatility in the interval $[t + 1, t + p]$:

$$\bar{r}^2_{t,p} := \sum_{i=1}^{p} r^2_{t+i} . \tag{34}$$

We calculated and compared the following MSE:

$$\mathrm{MSE}^*(p) := \sum_{t=1}^{n} (\bar{r}^{\,2}_{t,p} - \bar{\sigma}^{2,\,*}_{t,p})^2 \tag{35}$$

with * standing for SM or Garch. The MSE (35) is preferred to the simpler MSE

$$\sum_{t=1}^{n} (r^2_{t+p} - \sigma^{2,\,*}_{t+p})^2$$

because the latter uses a poor measure of the realized return volatility.[31] Through averaging, some of the idiosyncratic noise in the daily squared return data is canceled, yielding equation (34), a better measure against which to check the quality of the two forecasts.

TABLE 3.—COMPARISON OF VOLATILITY FORECASTING PERFORMANCE BETWEEN A GARCH(1,1) MODEL AND THE SM MODEL AT LONGER HORIZONS

Horizon p (days)	$\mathrm{MSE}^{\mathrm{SM}}(p)/\mathrm{MSE}^{\mathrm{Garch}}(p)$ S&P 500	$\mathrm{MSE}^{\mathrm{SM}}(p)/\mathrm{MSE}^{\mathrm{Garch}}(p)$ NASDAQ	Horizon p (days)	$\mathrm{MSE}^{\mathrm{SM}}(p)/\mathrm{MSE}^{\mathrm{Garch}}(p)$ S&P 500	$\mathrm{MSE}^{\mathrm{SM}}(p)/\mathrm{MSE}^{\mathrm{Garch}}(p)$ NASDAQ
10	0.26	0.94	110	0.51	0.60
20	0.50	0.93	120	0.51	0.58
30	0.41	0.83	130	0.53	0.56
40	0.44	0.83	140	0.54	0.54
50	0.42	0.78	150	0.56	0.52
60	0.44	0.76	160	0.56	0.49
70	0.44	0.71	170	0.58	0.47
80	0.46	0.70	180	0.59	0.44
90	0.46	0.66	190	0.61	0.43
100	0.48	0.64	200	0.62	0.41

The two models are reestimated every 20 observations using all the past observations. The ratio $\mathrm{MSE}^{\mathrm{SM}}(p)/\mathrm{MSE}^{\mathrm{Garch}}(p)$ is reported. A ratio smaller than 1 at horizon p indicates that the volatility forecast of the SM model at horizon p is more precise than that of the Garch(1,1) model. The figure shows overall better longer-horizon volatility forecast performance of the SM model.

Besides the S&P 500 data, the MSE analysis has been performed also on the returns of the NASDAQ index between October 12, 1984 and November 8, 2002. The data used in the out-of-sample variance forecasting comparison are the daily returns in the interval 1965–2000 (the data from 1957 to 1964 were used for initial estimation of the models) for the S&P 500 returns, and 1993–2002 for the NASDAQ returns (with the interval 1984–1992 for the initial estimation). The two models were reestimated every 20 days, and forecasts were made based on the most recent information available. Given the fact that its parameters are significantly varying through time, the Garch model, initially estimated on a sample of length 2000, was reestimated every 20 days on 2000 past observations.[32]

The results[33] of the comparison are given in table 3 and in figure 13. Table 3 reports the ratio[34]

$$\mathrm{MSE}^{\mathrm{SM}}(p)/\mathrm{MSE}^{\mathrm{Garch}}(p)$$

for $p = 10, 20, \ldots, 200$. Figure 13 gives more ample information on the ratio of MSEs. A ratio smaller than 1 at horizon p indicates that the volatility forecast of the SM

---

[30] If this condition is not fulfilled, the Garch(1,1) process, although possibly strongly stationary, has infinite variance.

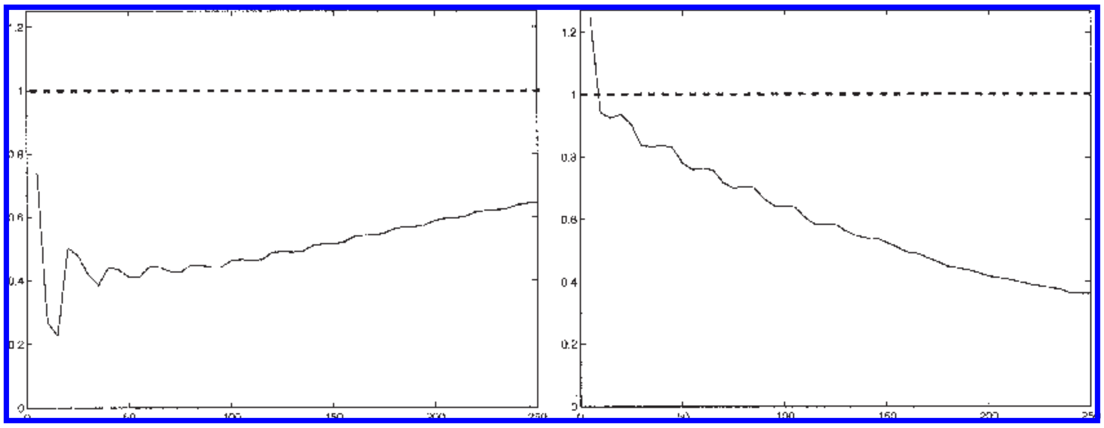
[31] It is well known (see Andersen & Bollerslev, 1998) that the realized squared returns are poor estimators of the day-by-day movements in volatility, as the idiosyncratic component of daily returns is large.

[32] The choice of the reestimation sample size of 2000 was based on the precision of the parameter estimation. In the case of a Garch(1,1) process, this sample size guarantees acceptable standard errors for the quasi-maximum estimators. See Straumann (2005).

[33] Using a misspecified maximum likelihood (Student t could be one of them) to estimate the Garch coefficients can yield inconsistent estimators (see Straumann, 2005). To guard against the possibly serious consequences of this type of misspecification, the Garch(1,1) coefficients were estimated using the quasi maximum likelihood method. The results of the forecasting exercise are qualitatively the same if a Student-t maximum likelihood is used for estimation of the model.

[34] The statistical significance of the results in table 3 has been assessed using the test of statistical significance for the differences of MSEs due to Diebold and Mariano (1995) (see also West, 1996; Harvey, Leybourne, & Newbold, 1997). The null hypothesis is that $Ed_{t,p} = 0$, where $d_{t,p} := \mathrm{SFE}^{\mathrm{Garch}}_{t,p} - \mathrm{SFE}^{\mathrm{SM}}_{t,p}$ is the loss-differential series, $\mathrm{SFE}^{\mathrm{Garch}}_{t,p} := (r^2_{t+p} - \sigma^{2,\mathrm{Garch}}_{t+p})^2$, and $\mathrm{SFE}^{\mathrm{SM}}_{t,p} := (r^2_{t+p} - \sigma^{2,\ \mathrm{SM}}_{t+p})^2$. The assumptions on the loss-differential series $d_{t,p}$ are those of stationarity and short memory. These assumptions are likely to hold when working under the null hypothesis of a stationary Student-t Garch(1,1) data-generating process. All differences in table 3 are significant at the 95% level.

FIGURE 13.—THE RATIO MSE^{SM} (p)/MSE^{GARCH} (p) FOR THE S&P 500 (LEFT) AND NASDAQ (RIGHT) RETURNS



On the x-axis, the horizon p. The Garch(1,1) model was reestimated every 20 observations using 2000 past observations. A ratio smaller than 1 at horizon p indicates that the volatility forecast of the SM model at horizon p is more precise than that of the Garch(1,1) model. The figure shows overall better longer-horizon volatility forecast performance of the SM model.

model at horizon p is more precise than that of the Garch(1,1) model.

The results in table 3 and figure 13 show superior volatility forecasting performance of the nonstationary model (18) to that of the Garch(1,1) model for the time series considered. They are consistent with the findings in Stărică (2003), where a Garch(1,1) process is fitted to the S&P 500 returns series between 1995 and 2003 using the quasi maximum likelihood. There it is shown that, in a setup close to the one described above, a simple locally constant volatility model closely related to equation (18) performed significantly better than the Garch(1,1) model in forecasting volatility over all horizons from 1 day to 1 business year. The same overall picture emerges from the detailed analysis of a large number of series of returns on various financial indices in Herzel et al. (2004).

The results presented in table 3 and figure 12 convincingly show that a stationary Garch(1,1) is a less appealing choice of data-generating process for the long series of daily returns on S&P 500 under discussion than the nonstationary model (18).

IX. Conclusions

In this paper an analysis of the S&P 500 absolute returns has been conducted giving up the usual assumption of global stationarity. We approximate the nonstationary data generating process locally by stationary models and identify the intervals on which stationary processes provide a good approximation. This is done using a goodness-of-fit test based on the integrated periodogram (Picard, 1985; Klüppelberg & Mikosch, 1996). Our approach leads to modeling the returns as a sequence of independent variables with a piecewise constant variance function. More concretely, the S&P 500 returns r_t can be described by the following:

$$r_t = h(t)^{1/2}\tilde{\epsilon}_t, \quad t = 0, 1, \ldots,$$

where $(\tilde{\epsilon}_t)$ is an i.i.d. sequence and $h(t)$ a function of time which can be well approximated by a step function, yielding a model with piecewise constant variance. We show that even a rough approximation of the variance dynamics by a step function is enough to explain most of the dependence structure present in the sample ACF of long absolute return series, providing an explanation for the so-called long-memory-in-volatility phenomenon.

We compared the forecasting implications of our nonstationary, unconditional modeling with, first, a stationary, long-memory paradigm, and second, a stationary, conditional methodology. Both comparisons show the superiority of the nonstationary, unconditional approach.

REFERENCES

Andersen, T., and T. Bollerslev, "Answering the Critics: Yes, ARCH Models Do Provide Good Volatility Forecasts," *International Economic Review* 39 (1998), 885–905.

Anderson, T. W., "Goodness of Fit Tests for Spectral Distributions," *Annals of Statistics* 21 (1993), 830–847.

Bartlett, M. S., Problèmes de l "Analyse Spectrale des Séries temporelles stationnaires, *Publications de l" Institut de Statistique de l' Université de Paris* III: 3 (1954), 119–134.

Brockwell, P. J., and R. A. Davis, *Time Series: Theory and Methods,* 2nd ed. (New York: Springer, 1991).

Cai, J., "A Markov Model of Unconditional Variance in ARCH," *Journal of Business and Economic Statistics* 12 (1994), 309–316.

Dahlhaus, R., "Fitting Time Series Models to Nonstationary Processes," *Annals of Statistics* 25 (1997), 1–37.

Diebold, F. X., "Modeling the Persistence of the Conditional Variances: A Comment," *Econometric Reviews* 5 (1986), 51–56.

Diebold, F. X., and A. Inoue, "Long Memory and Regime Switching," *Journal of Econometrics* 105 (2001), 131–159.

Diebold, F. X., and A. Mariano, "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics* 13 (1995), 253–263.

Drees, H., and C. Stărică, "A Simple Non-stationary Model for Stock Returns," working paper, www.math.chalmers.se/~starica (2002).

Giraitis, L., and R. Leipus, "Testing and Estimating in the Change-Point Problem of the Spectral Function," *Lithuanian Mathematical Transactions* 32 (1992), 20–38.

Granger, C. W. J., and N. Hyung, "Occasional Structural Breaks and Long-Memory," University of California, San Diego, discussion paper 99–14 (1999).

Granger, C. W., and R. Joyeux, "An Introduction to Long-Memory Time Series Models and Fractional Differencing," *Journal of Time Series Analysis* 1 (1980), 15–29.

Granger, C. W., and T. Teräsvirta, "A Simple Non-linear Time Series Model with Misleading Linear Properties," *Economics Letters* 62 (1999), 161–165.

Grenander, U., and M. Rosenblatt, *Statistical Analysis of Stationary Time Series,* 2nd ed. (New York: Chelsea, 1984).

Härdle, W., V. Spokoiny, and G. Teyssière, "Adaptive Estimation for a Time Inhomogeneous Stochastic Volatility Model," SFB discussion paper 2000/6, http://sfb.wiwi.hu-berlin.de (2000).

Harvey, D., S. Leybourne, and P. Newbold, "Testing the Equality of Prediction Mean Squared Errors," *International Journal of Forecasting* 13 (1997), 281–291.

Herzel, S., C. Stărică, and R. Tütüncü, "A Non-stationary Multivariate Model for Financial Returns," working paper, www.math.chalmers.se/~starica (2002).

Herzel, S., T. Nord, and C. Stărică, "Why Does the Garch (1, 1) Model Often Fail to Produce Reasonable Longer-Horizon Forecasts?" working paper, www.math.chalmers.se/~starica (2004).

Hidalgo, J., and P. M. Robinson, "Testing for Structural Change in a Long-Memory Environment," *Journal of Econometrics* 70 (1996), 159–174.

Hosking, J. R., "Fractional Differencing," *Biometrika* 68 (1981), 165–176.

Hsu, D. A., R. Miller, and D. Wichern, "On the Stable Paretian Behavior of Stock-Market Prices," *Journal of the American Statistical Association* 69 (1974), 108–113.

Klüppelberg, C., and T. Mikosch, "Gaussian Limit Fields for the Integrated Periodogram," *Annals of Applied Probability* 6 (1996), 969–991.

Lamoureux, C. G., and W. D. Lastrapes, "Persistence in Variance, Structural Change and the GARCH model," *Journal of Business and Economic Statistics* 8 (1990), 225–234.

Lobato, I. N., and N. E. Savin, "Real and Spurious Long-Memory Properties of Stock-Market Data," *Journal of Business & Economic Statistics* 16 (1998), 261–268.

Massey, F. J., Jr., "The Distribution of the Maximum Deviation between Two Sample Cumulative Step Functions," 22 (1951), 125–128.

——— "Distribution Table for the Deviation between Two Sample Cumulatives," *Annals of Mathematical Statistics* 23 (1952), 435–441.

Mikosch, T., "Periodogram Estimates from Heavy-Tailed Data" (pp. 241–258), in R. Adler, R. Feldman, and M. S. Taqqu (Eds.), *A Practical Guide to Heavy Tails: Statistical Techniques for Analysing Heavy-Tailed Distributions* (Boston: Birkhäuser, 1998).

Mikosch, T., and C. Stărică, "Long Range Dependence and ARCH Modeling" (pp. 439–460), in P. Doukhan, G. Oppenheim, and M. Taqqu (Eds.), *Theory and Applications of Long Range Dependence* (Boston Birkhäuser, 2003); www.math.chalmers.se/~starica.

——— "Non-stationarities in Financial Time Series, the Long Range Dependence and the IGARCH Effects," this REVIEW, 86 (2004), 378–390.

Nelsen, R. B., *An Introduction to Copulas* (New York: Springer, 1999).

Officer, R., "A Time Series Examination of the Market Factor of the New York Stock Exchange," University of Chicago PhD dissertation (1971).

Picard, D., "Testing and Estimating Change-Points in Time Series," *Advances in Applied Probability* 17 (1985), 841–867.

Priestley, M. B., *Spectral Analysis and Time Series,* vols. I, II (New York: Academic Press, 1981).

Shorack, G. R., and J. A. Wellner, *Empirical Processes with Applications to Statistics* (New York: 1986).

Simonato, J. G., "Estimation of GARCH Processes in the Presence of Structural Change," *Economic Letters* 40 (1992), 155–158.

Stărică, C., "Is Garch (1, 1) as good a model as the accolades of the Nobel prize would imply?" working paper, www.math.chalmers.se/~starica. (2003).

Stock, J., and M. Watson, "Evidence on Structural Instability in Macroeconomic Time Series Relations," *Journal of Business and Economic Statistics* 14 (1996), 11–30.

Straumann, D., "*Estimation in Conditionally Heteroscedastic Time Series Models* (Springer-Verlag, 2005).

Straumann, D., and T. Mikosch, "Quasi-MLE in Heteroscedastic Time Series: A Stochastic Recurrence Equations Approach," working paper, http://www.math.ku.dk/mikosch; *Annals of Statistics,* forthcoming (2005).

West, K. D., "Asymptotic Inference about Predictive Ability, *Econometrica* 64 (1996), 1067–1084.